VU Research Portal

Weed for Thought

Leonelli, S.

2007

document version
Publisher's PDF, also known as Version of record

Link to publication in VU Research Portal

citation for published version (APA)
Leonelli, S. (2007). *Weed for Thought: Using Arabidopsis thaliana to Understand Plant Biology*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl

Download date: 05. May. 2023

Weed for Thought
Using *Arabidopsis thaliana* to Understand Plant Biology

VRIJE UNIVERSITEIT

Weed for Thought
Using *Arabidopsis thaliana* to Understand Plant Biology

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Wijsbegeerte
op dinsdag 10 april 2007 om 15.45 uur
in het auditorium van de universiteit,
de Boelelaan 1105

door

Sabina Leonelli

geboren te Modena, Italië

promotor: prof.dr. J.A.Radder
copromotor: dr. H.W. de Regt

Voor Elisa

Table of Contents

This study was conducted with the support of the Netherlands Organisation for Scientific Research (NWO). The Netherlands Graduate School of Science, Technology and Modern Culture (WTMC), the Faculty of Philosophy of the Vrije Universiteit Amsterdam and the Casa Editrice Polifemo contributed financially to the publication of this volume.

Cover Images: Luca Leonelli, ink on paper (www.lucaleonelli.it)

Technical assistance: Michel Durinx and PrintPartners Ipskamp.

Acknowledgments

It is a great pleasure to thank at least some of the extraordinary people who accompanied me through my postgraduate experience. Thanks to them, the last five years have proved an incredibly intense period of growth and learning. Over and over again, my mentors and friends showed me what it means to be dedicated to research without ever forgetting one's humanity – an example that goes well beyond the ability to formulate a philosophical argument and that I hope to treasure for years to come.

As my supervisor and the director of the project 'Understanding Scientific Understanding', *Henk De Regt* provided me with invaluable support and advice, a rich philosophical and social framework in which to pursue my interests, as well as the possibility to share the intellectual journey he had started already at the end of his own PhD, with all its concerns and enthusiasms. *Hans Radder* showed me the meaning of the Dutch term 'promotor' by following, encouraging and challenging my every intellectual step – all without ever losing perspective and irony. Hans' and Henk's contributions stand out in every page of this thesis, each of which benefited from their careful thinking and illuminating critiques; the pleasure of discussing my work with them has been a major motivation for my research and writing efforts.

Hasok Chang has been a mentor and inspiration for my research since I first entered higher education in 1997. Without his support and encouragement I might have never pursued my studies at the postgraduate level; it is thanks to his advice that I applied for my PhD position at the Vrije Universiteit. It is impossible to properly acknowledge how much Hasok taught me over the years, ranging from a foundational understanding of philosophy, to research methods in history, to an appreciation for the human values, ideals and achievements characterising the fallible human enterprise that we call science.

Rasmus Winther, Rachel Ankeny and *Thomas Reydon* keep playing important roles in my formation in the philosophy of biology. This thesis owes much to our frequent exchanges - in particular, to Rachel's advice that I narrow my focus to Arabidopsis research. Conversations with *John Dupre*, *Michael Friedman*, *James Griesemer, Andrea Loettgers, Helen Longino, Martina Merz, Jan Willem Romijn, Edna Suarez, Henk van der Belt* and *Arno Wouters* greatly helped to shape the ideas expressed here. I am also grateful for the encouragement by members of the Konrad Lorentz Institute, especially *Werner Callebaut* and *Gerd Müller*; the support granted by ISHkabibble to attend its meetings; and the opportunity to visit the excellent Instituto de Filosofia of the UNAM in Mexico City.

At the Faculty of Philosophy of the Vrije Universiteit, I was lucky to share office, thoughts, ideas, meals, trips, parties, worries, meetings, coffees, [etc.] with *Kai Eigner* and *Nikki Smaniotto*. I also benefited from discussions with *Arianna Betti, Martijn 'ET' Blaauw, Lieven Decock, Peter Kirschenmann, Edwin Koster, Huib de Jong, Marije Martijn, Fritz Schipper* and *Mariette Willemsen.* The musical experiment called 'Otto e Mezzo' made me enjoy Amsterdam even more: bedankt lieve octetters! Y gracias *Eva,* para tu paciencia y entusiasmado apoyo durante esos anos difìciles.

My visits to Arabidopsis laboratories were crucial to my own understanding of research in plant biology. I deeply appreciate the time and energy invested by *Sue Rhee* and *Sean May*, as well as their staff respectively at the Carnegie Institute of Plant Biology and the Nottingham Arabidopsis Stock Centre, in helping me to study Arabidopsis research and allowing me access to their IT and archival resources. I am also grateful to *Maarten Koornneef*, *Chris Somerville* and *Shauna Somerville* for their availability and kindness.

As a member of the WTMC I benefited from excellent training in science and technology studies, including several workshops organised by *Paul Wouters*, *Annemiek Nelis* and *Sally Wyatt*. I owe warm thanks to them as well as to all WTMC PhD students with whom I spent days discussing countless aspects of STS research. Moreover, the Aio-Netwerk for the Philosophy of Science and Technology, as well as the informal research group ('DEMUS') based at the Vrije Universiteit, brought me in contact with the Flemish and the Dutch communities of philosophers of science. Exchanging ideas with *Maarten van Dijk*, *Jeroen van Bouwel* and *Marcel Boumans* has been especially helpful (and fun!). A third important source of support, learning, challenge and friendship has been the administrative board of the Graduate Journal of Social Science. I am deeply grateful to *Ward Rennen, Maud Radstake, Martijn Wit, Wouter-Jan Oosten, Esther Foreman, Stefanie Ortmann, Iris van der Tuin, Andre van Dokkum and Marcel Scheele* for the years of fruitful collaboration, good dinners and hard work towards bringing the journal to life.

The research group 'Scientific Practice and Epistemology' based at the Centre for the Philosophy of the Natural and Social Sciences at the London School of Economics guided my first steps in postgraduate research and offered me insight into the epistemic role of models and measurement. The LSE also brought me in touch with *Stefanie Ortmann, Gary Jones, Natasha Gilbert, John Gibler, Jacopo Cuscianna* and *Raffaele Marchetti*, all of whom contributed ideas and energy to my research efforts. Last but certainly not least, I thank my colleagues in the LSE project 'How Well Do Facts Travel?' for their support through last-minute work and worries towards submitting.

The interdisciplinary nature of this text shows the deep influence that my undergraduate tutors at the STS Department of University College London had on my thinking. Notably, *Joe Cain* infected me with his love for the life sciences, while *Brian Balmer* nurtured my penchant for the 'social' in STS. At UCL I also met *Emilia Sanabria*, whose friendship, intelligence and experience have inspired much of my work on embodied knowledge.

Finally, in line with my own argument, I hold my family and friends, with whom I share so much of my experience, to be highly responsible – though unfortunately not accountable – for my current ways of understanding the world. *Luca*, *Fany* and *Andrea* never failed to support my every endeavour and help me in every conceivable way (the least of which is my father's pictorial contribution to this book), while providing the most challenging example of intellectual dedication and love for life. No matter where they are, my friends know how close they are to my heart and deeds. Stefano: *grazie*.

Michel. *Sine sole sileo* (yet: *abusus non tollit usum*).

Chapter 1. Introduction: What Is the Grass?

A child said, What is the grass?
Fetching it to me with full hands;
How could I answer the child?
… I do not know what it is more than he
Walt Whitman, 1855

1.1 Scientific Explanations and Scientific Understanding

What is the grass? When posing this question in his ‘Song of Myself’, Whitman is not expecting an answer to be possible. He insists that none of his own experiences as a human being make him able to answer his innocent questioner. In his typically solipsistic fashion, Whitman interprets his existential quest, as well as the experiences involved in it, to hold for all mankind. Knowledge is elusive, he means to argue, no matter which actions we undertake, which culture we belong to and which methods we use to increase our understanding of nature.

There is a sense in which Whitman’s pessimism is easily dismissed. It suffices to survey the range and variety of cultural traditions and systems of knowledge that hitherto tried to make sense of the entities and processes characterising the mysterious environment in which humans evolve. This thesis considers one such system of knowledge, the one I am most familiar with and in which I have been educated: the understanding of nature provided by Western science. Scientists, particularly those among them specialising in plant biology, have been posing Whitman’s same question for centuries. In the course of time, the question has taken many forms and attempts to answer each of those developed into separate disciplines. ‘How does grass develop?’, asks the developmental biologist. ‘Why is grass green?’, wonders the physiologist. ‘How does grass grow and multiply?’, ponders the geneticist. ‘What does grass contribute to the environment?’, asks the ecologist. What is important to these ways of rephrasing the question is not just their increasing specialisation and the fact that they have been posed again and again by different groups of researchers. Thanks to the joint efforts of research groups armed with intellectual curiosity, sophisticated instruments and corroborated methodologies, biologists have actually found *answers* to those questions. For instance, the taxonomist lists several species of weed that might be defined as ‘grass’. The ecologist explains the function of grass in a given ecosystem, while the physiologist points to the structure of stem, roots and leaves and to the complex interaction of hormones that make processes such as germination possible. The molecular biologist zooms in on the microscopic mechanisms that allow plants to reproduce and develop. The evolutionary biologist tells a story about the ancestors of grass and how different species have been selected to keep replenishing our fields with green.

The accumulation of scientific explanations such as the ones I sketched, which together constitute a vast (and fast growing) body of knowledge about nature, is surely a great human achievement and one that has been examined and discussed within much philosophical literature. Yet, I intend to defend the idea that what makes scientific knowledge remarkable is *not* its role in explaining scientific phenomena. In this thesis, I focus on the extent to which scientific research allows us human beings to *comprehend* the phenomena that it aims to explain. Under which conditions do human beings learn to *use* scientific explanations *in order to* obtain understanding of natural phenomena? What does scientific understanding involve and how is it achieved? The main claim of this thesis is that scientific understanding is a necessary precondition for scientific knowledge: in other words, it is not acceptable for a person to state that she 'knows' a specific phenomenon (in a scientific manner), if this person is not able to use a scientific explanation to understand the phenomenon in question.

1.2 Scientific Experiences and The Ability to Understand

Why should one think of understanding as a precondition for scientific knowledge? After all, information abounds about all aspects of the natural world. Apart from books and encyclopaedias, internet search engines allow us to obtain scientific knowledge about virtually any subject in a matter of seconds. We can access and own scientific knowledge merely by entering a library or logging into a computer. Google the word 'grass' and you will obtain 107.000.000 entries[1], each of which constitutes a potential answer to Whitman's question. One would be tempted to classify Whitman's declaration of ignorance as the existentialist whining of a (possibly Luddite) poet.

In fact, I believe that Whitman's pessimism is relevant precisely in the context of such an ocean of information. When acknowledging his ignorance, Whitman is not declaring that knowledge about grass does not exist. He is stating that he is personally unable to make sense of that knowledge *in the light of his own experience*. Understanding a claim or explanation is a largely subjective matter. The subject trying to understand needs to employ skills, experiences and beliefs that are appropriate to achieving such a goal. Biologists represent, arguably, a case in point: their training, professional life and experimental work (if any) constitute experiences that are highly efficient in enabling them to understand a series of phenomena in certain specific ways. Whitman did not have such experience and the related skills: in this sense, he is certainly right in pointing to his inability to make sense of (especially scientific) knowledge concerning grass.

Following this line of thought, I shall base my analysis of scientific understanding on a study of contemporary research practices in the life sciences: that is, on my observations on how practicing scientists carry out research and share their results and methods with each other, thus acquiring the expertise that enables them to use scientific knowledge to understand natural phenomena. The study of contemporary biological research illustrates how skills and commitments acquired through specific practices inform the way in which biologists understand the world. I shall focus on scientific understanding as *a cognitive achievement derived through the adequate performance of specific activities,* which can lead to different interpretations of the same phenomenon *depending on the specific skills and commitments of the individual involved.* In my view, understanding is not an attribute

[1] Google access 27/01/2006.

of knowledge itself, which can be measured quantitatively (as in ‘how much do you understand?’). Rather, it is an ability that is acquired by individuals in a variety of different ways and that can therefore take different forms depending on the instruments that are used to exercise it. In the case of biological understanding, the quality of the ability to understand displayed by any individual researcher will depend on his or her acquisition of appropriate background knowledge *as well* as expertise in handling the instruments, models and theories that make it possible to produce and apply any scientific explanation. It also requires the social skills enabling effective interactions with other scientists, for instance in order to communicate one’s findings and assess their significance in the light of both personal experience and the standards adopted by the relevant scientific community. Indeed, access to knowledge does not automatically involve an awareness of how to interpret and apply such knowledge: this awareness needs to be built through years of training into specific modes of thinking and acting.

1.3 The Arabidopsis Research Community

An important corollary of my view is that there are many types of understanding, some of which can be classified as ‘scientific’ because they are developed within and adopted by one or more scientific communities. There is no intrinsic property that sets scientific understanding apart from other forms of understanding. This demarcation rather emerges from the specific practices, goals and theoretical background characterising the community in which scientific understanding is developed and performed. Scientific understanding is, thus, a subset of the many types of understanding of the natural world acquired by human beings, whose special characteristics are determined by the manner through which such understanding is achieved as well as the standards and tools through which it is diffused and evaluated.

This position implies that an investigation of *scientific* understanding should be at once analytic and empirical, since it has to define the features of scientific understanding as a separate and crucial notion to the study of science as well as to examine how such understanding grows out of the research practices of specific communities. I therefore aim, on the one hand, to offer an epistemological argument outlining which experiences and skills are most significant towards helping and informing a scientist’s understanding of the world. On the other hand, my analysis springs from a close investigation of the history, ethos and research practices characterising a specific community of plant biologists. This community, which I shall refer to as the ‘Arabidopsis community’, includes all scientists using the small flowering weed *Arabidopsis thaliana* as the material focus for their research and, in fact, as a representative of many other species of plants, including all other types of grass.[2]

With over 6000 laboratories spread across all five continents devoted to its study, Arabidopsis is the most widely used model organism in contemporary plant biology. In line with the rest of the life sciences, plant biology is a messy affair involving a plethora

[2] I shall henceforth refer to *Arabidopsis thaliana* as Arabidopsis, thus, for the sake of convenience, doing away with the italics used to characterise its complete name as the name of a species.

of different research groups with varying interests and expertise, each of which investigates different aspects of plant structure, evolution and development. The Arabidopsis community is no exception: it includes widely diverse fields of inquiry and expertise, each of which specialises in a different aspect of Arabidopsis biology. This is a landscape of multiple worlds, each of which characterised by its own tools, concepts, goals, theories and ethos, not to mention the different nationalities and cultural contexts of the thousands of researchers involved. It is therefore very difficult to view the landscape in just one glance: one has to switch disciplinary glasses to be able to acknowledge the richness and diversity of the research being done.

One substantial challenge, given this fragmentation, is finding ways to systematise and integrate such knowledge. This is where the choice of the Arabidopsis community, rather than any other groups of biologists, becomes especially interesting to the study of scientific understanding in the life sciences. Despite its fragmentation into independent and diverse units, the community has a highly centralised structure. The coordination among component groups and the choice of research directions are still largely influenced by the handful of powerful, charismatic scientists who started Arabidopsis research in the early 1980s. Since that time, these individuals have been encouraging their pupils and associates to adopt a cooperative ethos in their research practices – a choice supported by the belief that only a joint effort would enable them to revive and re-invent plant science as a whole (mostly, as we shall see, by giving it a strong foundation at the molecular level). A notable consequence of this situation is that the research carried out in the Arabidopsis community is strongly centred on two goals:

(1) the goal of *integration*, that is of organising all the results gained about the different aspects of Arabidopsis into a unique body of knowledge, thus encouraging the understanding of the plant as a whole;

(2) the goal of *representativeness*, i.e. of treating Arabidopsis as a model for any plant, rather than as a model of a specific type of weed.

This last goal implies treating knowledge about Arabidopsis as potentially extending to all plants, including all types of weed and grass. This characterisation has obvious advantages, as it allows scientists to gain an immense amount of widely applicable knowledge by focusing all research resources on one organism. Further, using a single organism enormously facilitates the achievement of the first goal, the goal of integration, as it allows to study the relation among different aspects of plant biology without worrying about inter-species variation (a method widely supported in the emerging field of 'system biology'). In other words, the concentration of research efforts on a single type of weed might be more effective than a comparative study of several different species in order to obtain an integrated, comprehensive answer to broad questions about plants (such as Whitman's question about grass). These advantages notwithstanding, many biologists see these research strategies and goals as highly suspicious and potentially damaging to biology as a whole, since it is not clear to which extent Arabidopsis can be taken to represent other plants – and anyhow, assessing the representativeness of Arabidopsis would imply studying other plant species just as thoroughly. These critiques make the second goal of the Arabidopsis community look untenable in the light of the huge diversity of life forms populating our planet.

It is presently impossible to determine the role that Arabidopsis research will assume within future biology: it might be that its goals and results will prove to be much less foundational than is currently thought. Even in the light of this uncertainty, which anyhow characterises all scientific projects at least to some extent, I see Arabidopsis research as a valuable locus for examining how biologists understand phenomena and exchange such understanding among each other. In particular, I shall centre my attention on two highly powerful sub-groups within the community, whose efforts are directed towards constructing conceptual and material tools that might facilitate the development of an *integrated approach* to the study of Arabidopsis biology. These are The Arabidopsis Information Resource [TAIR] hosted by the Carnegie Institute for Plant Biology in Stanford, CA and the Nottingham Arabidopsis Stock Centre [NASC] based at the University of Nottingham, UK. In the context of an ever-fragmented and pluralistic field such as biology, the task pursued by these sub-groups proves as difficult as it is ambitious. It requires them to critically consider questions such as 'what is biological knowledge' and 'how do we understand it', so as to construct tools that would help other biologists towards understanding not only some bits of knowledge about Arabidopsis, but also the relations among these bits. Biologists working in these groups are not only unhindered, but indeed encouraged to ask biological questions in a very broad, inclusive manner. Because of this width of scope, as well as the emphasis on producing tools for displaying and understanding knowledge, this type of research represents an excellent ground for investigating which factors facilitate the scientific understanding of Arabidopsis biology (both by Arabidopsis scientists and by other biologists).

1.4 Preview of Thesis Content

Ultimately, this thesis is about scientific investigations and learning as experienced by individual researchers and shared within scientific communities. Understanding is defined as a cognitive achievement whose emergence and characteristics depend on the ensemble of skills and commitments acquired by each individual scientist through three main types of experience: *intellectual*, including reasoning through and developing available concepts, theories and explanation for natural phenomena; *material*, involving the setting and manner of intervention by researchers in the phenomena under scrutiny; and *social*, as implied by the dependence on and contribution of individual scientists to one of more scientific communities sharing the same research interests. The social contexts within which such understanding is developed is particularly important to its definition as 'scientific', since the personal experience of understanding is transformed in communicable, socially shared experience by virtue of the individual's participation in one or more scientific communities. I hope to be able to clarify the motivations and implications of this view of scientific understanding through a close historical and sociological study of some of the practices characterising Arabidopsis research. This proves to be a complex endeavour, requiring me to weave together an extensive web of philosophical, historical and sociological analysis. Before delving into the material, I will therefore briefly outline the structure of this dissertation, which is organised in steps of increasing complexity: each new chapter builds upon the previous ones, thus progressively disclosing various aspects of my argument and allowing me to discuss their philosophical significance and implications in connection to the concrete features and processes characterising Arabidopsis research.

A major goal for any philosophical project consists of clarifying the background, content and methodology used to pursue a given question. Chapters 2, 3 and 4 are thus devoted to introduce and justify the philosophical, historical and methodological bases for my argument. *Chapter 2* sets the philosophical stage for my analysis. I start by reviewing the contemporary philosophical debate on the difference between explanation and understanding. I then zoom in on the biological sciences and discuss the characteristics of biology that I deem to be of relevance to the study of understanding, such as the pluralism characterising both theories and models and what I call the 'dual nature' of biological knowledge. In this context, I provide a first definition of scientific understanding, which I exemplify and elaborate throughout chapters 5,6 and 7 and to which I return to in a systematic fashion in Chapter 8. I conclude the chapter by exposing and discussing the elements of social epistemology that are relevant to my investigation of how individual understanding is formed through participation in one or more scientific communities. In *Chapter 3*, I outline the historical and social context of my research. After a short discussion of 'model organism communities' in biology, I trace the origins and development of the community working on Arabidopsis. I then discuss the current structure and goals of the community and I examine in detail the social and scientific roles played by the TAIR and NASC research groups. *Chapter 4* aims to present and defend my research methodology, which I classify under the heading 'history, philosophy and social studies of science and technology', in the light of the philosophical and historical elements provided in the previous chapters. Starting from a discussion of the significance of using case studies to investigate broad philosophical issues, I reflect on the motives and concerns underlying my research and connect those with the research methods that I adopted. As I explain in this chapter, it is my hope that this piece of research will be of interest to both philosophers and historians of science, as well as to philosophically inclined biologists (such as the many scientists that have helped me by granting access to their laboratories, personnel and thoughts on Arabidopsis research).

Chapters 5, 6 and 7 constitute the main argumentative body of the thesis. Each of them focuses on a different aspect of biological understanding, as exemplified in the context of Arabidopsis research. Together, Chapters 5 and 6 constitute an extended reflection on the role played by two types of epistemic skills, which I call theoretical and performative, towards enabling biologists to understand phenomena. *Chapter 5* focuses on research practices in TAIR, which consist of producing databases able to visualise as many of the available data hitherto gathered on Arabidopsis as possible. I show that the construction, as well as the use, of such databases involves reference to a theoretical framework, which in this case consists in the network of concepts ('gene ontology') used by TAIR researcher to order and classify Arabidopsis data. This leads me to examine the skills needed by Arabidopsis researchers in order to use these data to understand Arabidopsis biology. After tracing the distinction between theoretical and performative skills, I highlight how biologists exercising different sets of skills in their research might acquire different understandings of the same phenomena. This claim is developed in *Chapter 6*, where I discuss the activities carried out by NASC researchers in order to produce Arabidopsis specimens for distribution and use by laboratories around the world. I argue that both TAIR and NASC successfully produce models of Arabidopsis: in the former case, the models consist of images of the biology of the plant; in the latter case, the models consist of actual organisms, whose features have been selected and modified to fit the demands of experimenters. A comparison between the production and use of these two types of models brings out significant differences in the skills required to manipulate them. Through an analysis of the way in which these models are abstracted from, respectively, Arabidopsis data sets and actual Arabidopsis wildtypes, I elaborate on the manner in which theoretical and performative skills can be combined in order to use these models to enhance one's understanding of Arabidopsis biology. *Chapter 7* adds a crucial layer of complexity to the analysis developed in the previous chapters, by reviewing the social dynamics surrounding Arabidopsis research. I characterise the Arabidopsis community as a case of centralised big science and I analyse the effect of this institutional arrangement on the acquisition and distribution of scientific understanding within the community. In particular, I list several social skills that I deem to be necessary for scientists to be able to understand Arabidopsis biology, insofar as they grant access to the theoretical and performative skills required to conduct as well as share research with other biologists.

Finally, *Chapter 8* presents the epistemological conclusions drawn from my analysis of how skilful recourse to theories, models and socially shared resources enables Arabidopsis researchers to enhance their understanding of plant biology. I point to the strong ties between researchers' skills and the commitments they make to the actions and beliefs through which such skills can be exercised. Different kinds of biological understanding of a phenomenon can be obtained depending on which combination of skills and commitments is used by researchers to coordinate their theoretical and embodied knowledge of that phenomenon. I sketch some implications of this pluralistic vision of understanding in biology, while also briefly pointing to the research paths opened by the approach herein proposed.

Chapter 2. Investigating Understanding in Biology

> *Whatever we mean by understanding (at least as a scientific goal), it is a safe presumption that only humans can do it.*
> *Evelyn Fox Keller 2002, 322*

2.1 Understanding and Explanation in the Philosophy of Science

2.1.1 The Long Way From the Received View

Rarely mentioned and scarcely reflected upon, scientific understanding seems to have been largely dismissed as a topic worth of inquiry by 20th century analytic philosophers, at least until very recently. This century-long dislike is epitomised by the work of Carl Hempel, whose thoughts on the issue were most forcefully expressed in the volume *Aspects of Scientific Explanation*, published in 1965. Hempel's views were so influential as to deserve the label of 'received view' on explanation in science (proposed, somewhat polemically, by Putnam [1962]). In Hempel's eyes, understanding is undeniably an important pragmatic feature of scientific research. However, it is only relevant to scientific epistemology insofar as it is linked to a deductive-nomological type of explanation. 'The explanation enables us to understand why the phenomenon occurred' (Hempel 1965, 337). Understanding is an inseparable by-product of any good scientific explanation. According to Hempel, it is therefore on explanation, rather than understanding, that philosophical inquiry should focus.

This thesis questions the idea that scientific explanations hold the key to unravelling the epistemic role and significance of understanding in science. Decades of philosophical work on the notion of explanation did not provide us any tools to think about understanding, that is, about *what it means* to understand why or how a given phenomenon occurs, *regardless of* – or, only partially depending upon - the actual answer given to that 'why' or 'how' (that is, regardless of the contents of the explanation itself). Hempel argued that this issue is of no interest to philosophy of science. This is because it involves studying the characteristics of the *subject* acquiring understanding (the scientist) just as carefully as the characteristics of the *object* that is understood (the natural phenomenon). In other words, it requires the philosopher to investigate the actual conditions under which scientific understanding is acquired. The pragmatic character of this approach inevitably leads to context-dependent results, since the conditions under which understanding is obtained may vary depending on what is understood, who understands and how. This variability makes it impossible, at least on purely rational grounds, to formalise universally valid conditions for understanding. Furthermore, Hempel notes that there is nothing to set the notion of scientific understanding apart from other types of human understanding (1965, 413). To him, understanding a sketch by Monty Python functions in exactly the same way as understanding Newtonian mechanics: it is the nature of what is understood that makes the difference between understanding a joke and a scientific law. The problem of defining the nature and epistemic role of understanding is therefore irrelevant to an epistemology of science and should be confronted by general philosophers or, even better, by cognitive scientists.

The Hempelian stance on understanding underscores a normative stance that was enforced by most Anglo-American philosophy of science developed between the 1950s and the early 1970s. This is the idea that the main goal of philosophical inquiries about science should be the formalisation of *universally valid principles* for the production and evaluation of *objective* knowledge about natural phenomena. This formalisation would be largely a rationalist enterprise – that is, it would be based on logical, a priori reasoning about the best ways in which valid and truthful knowledge can be gained. Granted this goal, Hempel's arguments are extremely forceful: to this day, his work represents a most valuable defence of traditionally objectivist methodology. The focus and commitments of the philosophical community have, however, shifted considerably in the last three decades. Several scholars have come to agree that the formulation and appraisal of norms applying to scientific research, which is arguably an inevitable part of any philosophical discussion of science, should be informed by *descriptive* knowledge of what scientists actually do. The caricature of the scientist as a knowledge-gathering robot, mindlessly following universal rules, is hardly an adequate description of scientific practices. An alternative is to consider the scientist as an active and fallible agent, endowed with limited skills and faltering motivation, who is responsible for making choices and forming judgements according to specific circumstances and inputs. The philosophical research agenda is increasingly shifting towards a throughout exploration of the implications of this insight for the nature of scientific knowledge and the objectivity and authority typically bestowed upon it.

Three main intellectual sources have been instrumental to supporting this approach. Taking inspiration from the work of Thomas Kuhn, *historians of science* have gathered an immense body of detailed knowledge about actual processes of discovery and theory-change and about the relation among material, social and theoretical components of scientific endeavours. This work encouraged reflection on the relevance of the individual scientist's personality, social and political context and economic interests to the way in which she gathers data, sets up experiments and interprets results. Titles of influential books such as Jan Golinski's *Making Natural Knowledge* (1998) are highly suggestive of the increasing historical interest in the heterogeneous processes by which science is constructed and disseminated.[3] *Sociologists* also contributed heavily to this empirical turn in philosophy by examining the structure and organisation of scientific communities and their relation to political and social institutions, media culture and the wider public. Research in this area spans from ethnographic investigations of scientific practices within small groups (for instance, in one laboratory, as in Latour and Woolgar 1979) to theoretical frameworks interpreting the interaction of both human and non-human actors in broad networks, of which scientists constitute but a subset (e.g. Bruno Latour's actor-network theory, as in his 1987). Despite accusations of being too steeped in the constructivist tradition (see Chapter 4 on this), the material gathered by social studies of science and technology marked a watershed in contemporary reasoning about how science works.

[3] Influential work in this respect has been carried out by, among others, Fox (1971), Shapin and Schaffer (1989), Jardine (2000), Daston (2000), Mendelssohn (1984, 1976), Greene (1976), Griesemer (1990, 2006), Rheinberger (1997), Chang (2004) and de Regt (1999, 2001).

The third field that proved most helpful to a re-examination of scientific research is the one of *gender studies*, particularly the epistemological reflections elaborated by feminist scholars in order to critique traditional objectivist (and, in their eyes, characteristically 'male') cosmology. This research yielded basic resources in order to explore the role of subjects in the production and validation of knowledge, three of which are especially relevant to the study of science. The first is Donna Haraway's work on technoculture, resulting in a view of knowledge as necessarily *situated* and context-dependent (Haraway 1988).[4] This approach, criticised for its outright rejection of the belief in science as an objective, authoritative source of knowledge, is nicely complemented by Sandra Harding's reflection on the difference between the notion of objectivity denoting context-independence and what she calls *strong objectivity* (Harding, 1986).[5] Both the notion of situated knowledge and that of strong objectivity require a throughout re-examination of the role of values in science, following the spirit of Kuhn's preliminary list of epistemic values that are relevant to theory-choice (Kuhn 1977, 331). A lasting contribution in this sense is Helen Longino's analysis of values that are *constitutive* of scientific knowledge, that is, that are 'the source of the rules determining what constitutes acceptable scientific practice or scientific method' (1990, 4). Longino distinguishes these constitutive values from contextual values that are also epistemologically relevant, but which pertain to 'the social and cultural environment in which science is done', rather than constituting reference points in scientific research (ibid.).

These advances in the historical and social analyses of science provided a new awareness of the extent to which personal judgement, skills, experiences and alliances foster, rather than hamper, scientific understanding. Science is undeniably a human endeavour. Its representation as an objective body of knowledge is both misleading and delusional. Far from securing access to a reliable and cumulative body of knowledge, traditional reliance on the de-personified objectivity of scientific theories masks the unavoidable situatedness and context-dependence of the observations on which those theories are based. In order to understand scientific understanding, extended reflection is needed on the role played by 'subjective', individual experience in enabling and shaping the concerns, perspectives and expertise of which scientific theories are but one incomplete expression.

Philosophers of science have risen to this challenge by paying closer attention to scientific practices as a complex and composite set of activities conducted by specific individuals in a particular institutional and social setting. Much has been published on the differences among practices in the special sciences, for instance in the procedures favoured in each science and the resulting variation in the types of theories, models, instruments and data that are produced and used. Most interestingly for my purposes here, variations among types of explanation have been subject to much debate.[6] The

---

[4] A similar emphasis on the locality of knowledge can be found in Rouse (1987, chapter 4).
[5] According to Harding (1986), situated scientific knowledge needs to point explicitly to the context (including personal interests) in which it was gathered. Harding highlights that there need not be a priori demarcations determining which contexts and interests are scientifically adequate: rather, awareness of the motivations behind a scientific view makes it easier to fruitfully apply and assess it, thus increasing its scientific value as a 'strongly objective' position.
[6] The main types of explanations subject to philosophical debate are causal explanations (Salmon 1990, 1998); mechanistic explanations (Craver 2001, Machamer et al 2000, Woodward 2003); functional

resulting consensus seems to be that there are many types of explanations (such as the causal, functional, unificationist or mechanistic), each of which provides meaningful knowledge and none of which is reducible to another type of explanation without significant loss of information. Science necessarily involves explanatory pluralism, a result that stands in opposition to Hempel's account of scientific explanation as unified and universal[7] and that emerges from a close examination of how and why explanations are actually produced (more on this in section 2.2.1).

2.1.2 Contemporary Attitudes to the Problem of Understanding

Given these developments, it may come as a surprise that most contributions to the philosophy of explanation still abide to the received view when it comes to discussing the notion of understanding. The tide might now be changing, with some systematic analyses dedicated to this subject appearing in print in the last three to four years[8] and as demonstrated by this very dissertation. Apart from these scattered pronouncements, however, the majority of philosophers of science is showing a disappointing reluctance in exploring the issue. I distinguish four main types of reaction to the idea of a philosophical examination of scientific understanding, each of which is grounded on a different perception of the relation between understanding and explanation.

a. *Indifference*. Many philosophers have taken quite literally Hempel's stance on the irrelevance of understanding to philosophy of science. They mention this term without wasting time on attempts to define its meaning independently from explanation. This holds irrespectively of the many critiques that those same philosophers successfully level against other aspects of Hempel's work. In his influential review 'Four Decades of Scientific Explanation', for instance, Wesley Salmon carefully exposes reactions and critiques of Hempel's deductive-nomological model of explanation, while not paying any attention to the notion of understanding if not as a self-evident outcome of explanation (Salmon, 1990).

b. *Reduction to Explanation.* Philosophers in this category also consider understanding to be a by-product of explanation, even if not agreeing on Hempel's arguments for this claim. An example is Michael Friedman's view on explanatory unification (1974). Friedman recognises the difference between the notions of understanding and explanation, insofar as he defines understanding as an epistemic goal secured by unifying explanations. His conclusion in the light of this definition is, however, that it is explanation, and not understanding, that deserves philosophical interest.[9] Philip Kitcher (1981) and Stephen Toulmin

---

explanations (a review of the main relevant literature is given by Wouters, 2005); and unificationist or subsumptive (Friedman 1974, Kitcher 1981; critique by Barnes, 1992).

[7] See Rose (1997), Beatty (1997), Keller (2000) and Dupré (1993) on explanatory pluralism.

[8] De Regt (2001), (2004); de Regt and Dieks (2005); Yi (2002).

[9] In fact, Friedman (1974) proposes a definition of understanding in science that depends entirely on the number of assumptions used in the explanations derived from that understanding. I should note here, however, that Friedman (2001, *pers. comm.)* has rejected this approach and currently proposes a very

(1961), despite their radically different interests and approaches, reach similar conclusions.[10]

c. *Hostility*. This group of philosophers shares the Hempelian intuition that a philosophical analysis of scientific understanding is not only futile, but also confusing and possibly harmful to the development of philosophy of science. Trout (2002, 2005), for instance, advocates a position similar to the Hempelian one by putatively demonstrating that the subjective features of understanding are of no concern to philosophy but are a matter of study for cognitive scientists. The differences between explanation and understanding, if any, should then be accounted for at the subjective level of individual psychology. More belligerently, Trout also maintains that all attempts to investigate them philosophically will result in a distorted view of scientific method (for a critique, see de Regt 2004).

d. *Non-committed Interest*. Authors such as Cushing (1994) and Keller (1999, 2002) do express a strong interest in scientific understanding as a promising research topic. Nevertheless, they stay clear of general pronouncements on this notion, again because of its vague and pragmatic nature. James Cushing maintains that understanding should be wedded to the ontological narrative underlying each particular theory. He does not, however, elaborate on how this conception differs from the role of explanation and indeed he does not provide any clear definition. Even more striking is Keller's study of explanatory strategies employed in 19th and 20th century experimental biology. There, she concludes that abstinence from analytic pronouncements concerning the nature of understanding and its relation to explanation is inevitable: 'understanding is a notoriously unstable word, and a central aim of this book has been to demonstrate this instability' (Keller 2002, 296); thus she will not try to define such notion 'for obvious reasons' (1999, 322).

This dissertation is intended as a critique of all four of these views. Rather than tackling the claims of each group in turn, however, I will respond to their arguments by proposing what I consider to be an epistemologically rich, alternative way of thinking about scientific understanding. Presently, I would like to take issue only with the last type of reactions, and particularly with Keller's argument for the inevitable instability – and thus, unintelligibility - of the notion of understanding.

There are two main reasons for viewing Keller's disinterest in a systematic analysis of understanding in science (what it is, how it is obtained and with which implications) as unjustified. First, her account is built on a number of assumptions concerning the

---

different view on the role and achievement of understanding within a scientific community. His earlier position still constitutes an important reference to many unificationists (e.g. to Kitcher), but not to himself.

[10] Toulmin's position might seem far from reducing understanding to explanation, as it focuses explicitly on intelligibility and on the importance of understanding in science. Yet, he adopts a tendentiously static view of the role played by standards of intelligibility in different phases of the history of science, thus paving the way for a Kuhnian understanding of scientific theory-change. His 'ideals of natural order' account for the achievement of scientific understanding provided that, in a somewhat circular fashion, we define understanding itself as the ensemble of theoretical presuppositions (indeed, 'ideals') by which we justify the results obtained at every step of our investigations (1961, 34).

meaning and applications of the notion of scientific understanding. Keller's refusal to make such assumptions explicit implies conferring an unqualified ubiquity of meaning to the word 'understanding', which in turn means a loss of analytic strength for her otherwise convincing account. For instance, her position on the relation between the two key notions of explanation and understanding (both of which are used in the titles of her book and of its final chapter) remains unclear. On one hand, she often treats understanding and explanation as distinct notions.[11] On other occasions, understanding is treated as synonymous not only with explanation, but indeed with any other term referring to the gathering and development of knowledge (including 'learning' and 'theorising'; Keller 2002, 2, 117 and 296). As a result of this conceptual vagueness, her position on what counts as explanation itself keeps shifting, making it very difficult to establish the import of her conclusions to general philosophy of science. Second, I share Keller's conviction that scientific understanding is likely to come in a variety of forms depending on the specific context and goals of the science involved. But is this not the case with scientific explanation, too? After all, the notion of explanation has been agreed to encompass a plurality of types, a taxonomy of which can be traced by looking at how explanations are produced and used in practice - as beautifully demonstrated by Keller's own work. Why, then, should diversity and context-dependence make us refrain from dissecting the analytic differences among types of understanding? The acknowledgment of the pluralism and instability of scientific practices is a constructive starting point for an investigation of understanding, in the same way as the recognition of explanatory pluralism increases, rather than diminishes, the value of studying different types of explanation.

Keller fails to engage in two philosophical tasks that I maintain to be necessary to a credible analysis of scientific understanding. The first of such tasks is the provision of an analytic framework in which this notion can be studied. This framework does not have to imply a strict definition of understanding, but rather provide philosophers and scientists with tools with which to assess the different ways in which understanding can be achieved and the forms that it can take. Such a framework is indispensable to highlighting the epistemological and scientific relevance of different ways of pursuing understanding in science, as well as different conditions under which understanding can be achieved. Its development would also allow to pursue the second task, that is, to provide a clear account of the way in which the notion of understanding differs from the one of explanation. This thesis aims to engage with these two tasks, thereby contributing to the birth of a philosophical debate on the notion of understanding that is based on the study of actual scientific practices. In Chapter 4, I shall discuss the methodological challenges posed by studying scientific understanding through a philosophy of scientific practice. The rest of the present chapter focuses on the epistemological framework within which I construct the claims on biological understanding presented in the conclusion of this thesis.

---

[11] For example when claiming that 'the need for understanding, as for explanation, is a human need, and one that can be satisfied only within the constraints that human inquirers bring with them' (Keller 2002, 297).

2.2 Focusing on Biology

I purposefully limit my study of scientific understanding to biology - in particular, as I illustrate in detail in Chapter 3, to the segment of plant sciences using *Arabidopsis thaliana* as model organism. Partly, this choice stems from my own fascination and familiarity with the 'sciences of life'. It also underscores my interest in exploring how the types of knowledge and practices that characterise biological research contribute to biologists' understanding of the phenomena that they study.[12]

My approach to the study of scientific understanding is restricted to the analysis of a well-defined set of practices. In this sense I follow the idea put forward by Henk de Regt (2001, 2005): an analysis of scientific understanding should 'encompass the historical variation of specific intelligibility standards employed in scientific practice' (de Regt and Dieks 2005, 138). De Regt goes on to propose a generally applicable definition of understanding as an epistemic *aim* of science, (ibid.,139). By contrast, I wish to propose an analysis grounded in a detailed investigation of specific features of research carried out in contemporary biology. I shall therefore not focus on the notion of understanding as a general (what de Regt and Dieks call 'macro-level': 2005, 140) aim of science, but rather I shall emphasise a reading of scientific understanding as a cognitive achievement shaped by scientists' experience and training within specific research communities. My interest lies with the processes and conditions through which such understanding-as-aim is achieved: in other words, in the activities underlying scientific understanding 'in action'. The results of this approach are compatible with de Regt's ideas and I shall specify precisely how the two accounts complement each other in the final chapter of this thesis (section 8.1.2).

In this section, I lay out the epistemological framework for my approach by depicting my view on biological knowledge and the relation among theories, models and phenomena that such a view implies. In particular, I emphasise the role of human agency in creating, applying and interpreting biological knowledge. My philosophical journey departs, again, from the Hempelian account of scientific understanding. In the previous section, I pointed to the rationalistic and universalising character of that view as two problematic factors, in the light of recent scholarship reflecting on science as it is actually practiced in local settings. I then sketched how the philosophical study of scientific understanding is affected by the clash between practice-inspired and rationalistic accounts of science. I now intend to discuss in more depth the account of the nature of scientific knowledge that is implied by the Hempelian view.

Hempel's work is strongly characterised by two interrelated assumptions about the nature of scientific knowledge. One is the idea that all scientific explanations, no matter their scope or generality, contain, at least implicitly, reference to *law-based generalisations*

[12] As I discuss in more detail in Chapter 4, I do not intend my reflections to be necessarily restricted to the biological sciences. It might be argued that other sciences, both natural and social, display characteristics similar to the ones that I attribute to biology, and thus that my account of how understanding is achieved is applicable to them, too. The scope of applicability of my analysis, as well as its validity, can only be determined by further scrutiny and research.

about natural phenomena that aim at the *unification* of scientific knowledge about the world. I come back to this assumption below. The second assumption is that explanations, observations and theories (that is, the main components of what he views as scientific knowledge) should all be expressed *propositionally,* that is, through a series of statements. Hempel recognises that scientific knowledge encompasses statements of different types. He thus distinguishes between observational and law-like statements. *Observational statements* are statements describing a specific phenomenon at the lowest level of generality. Hempel defines them as 'a sentence – no matter whether true or false – which asserts or denies that a specified object, or group of objects, of macroscopic size has a particular observable characteristic, i.e. a characteristic whose presence or absence can, under favourable circumstances, be ascertained by direct observation' (1965, 102). For instance, the sentence 'this group of storks is red-legged' counts as an observation sentence. *Law-like statements* spell out concepts (and relations among those concepts) that are applicable to more than one observation statement (ibid., 345): an example is the sentence 'all storks are red-legged' (ibid., 105). These law-like statements are also referred to as 'covering laws', precisely to highlight their role in unifying various observation statements under a unique, general explanation.[13]

When making the distinction between observational and law-like statements, Hempel is careful to underline that both types of sentences constitute crucial components of scientific knowledge. He is therefore far from advocating that all observational statements should be, at least in principle, reducible to law-like statements. At the same time, however, Hempel suggests that scientists should strive to find law-like statements that can be used to explain the observational statements gathered during empirical (experimental or field) research. This is because of the special role that law-like sentences play in providing what he calls 'objective insight' in phenomena:

> What scientific explanation, especially theoretical explanation, aims at is not [an] intuitive and highly subjective kind of understanding, but an objective kind of insight that is achieved by a systematic unification, by exhibiting the phenomena as manifestations of common underlying structures and processes that conform to specific, testable basic principles (Hempel 1966, 83).

As argued by Philip Kitcher (1981, 508), paragraphs such as the above reveal two basic intuitions underlying Hempel's view: (1) the belief that law-like statements constitute the most significant theoretical result of scientific research, insofar as they provide a foundation for the explanations through which we understand the natural world; and (2) the idea that the best way to magnify the explanatory power of these explanations is to

[13] The following quote from Suppe is useful in trying to reconstruct Hempel's view on scientific theories: 'Of the various concepts occurring in the pre-axiomatic version of the theory, a small number of these concepts are selected as basic; axioms are introduced which specify the most fundamental relations holding between these basic concepts; and definitions are given specifying the remaining concepts of the theory in terms of these basic ones. The relations specified by the axioms and definitions do not explicitly state the entire content of the theory, but if the axiomatisation is fruitful and adequate, it will be possible to deduce the remaining content of the theory from the axioms and definitions by a process of logical manipulation' (1977, 64). What Suppe here calls 'axioms and definitions' are the nomic assumptions – the 'covering laws' – on which Hempel constructs his generalisations.

subsume all available observational statements under the smallest possible number of law-like statements. These are the assumptions that allow Hempel to define the acquisition of scientific understanding as a by-product of the acquisition of scientific knowledge: a phenomenon is understood once law-like statements are formulated that might be employed to explain it.[14]

I intend to criticise Hempel's simplistic connection between 'possessing knowledge' and 'understanding phenomena'. In this section, I use recent studies of contemporary biological research to illustrate how neither of Hempel's assumptions on the nature of scientific knowledge fits actual scientific practice, its results and aims.[15] I first examine the theoretical results that biologists actually produce in their research activities. Both within and across different biological disciplines, several different types of theories are produced, some of which do not even contain law-like generalisations (let alone attempt to unify observations under a unique theoretical framework).[16] Far from seeing it as a problem to their research, biologists actually appreciate this pluralism in types of biological theories, which is tightly intertwined and sustained by a wide diversity among the modeling practices that are used to pursue and apply each type of theoretical result. Those modeling practices, in turn, include several ways of representing theoretical knowledge that are not propositional. Statements are but one manner to express biological knowledge, and not necessarily the most popular. Pictorial, three-dimensional, symbolic, digital models are equally if not more widely used in order to represent theoretical results – and, what is worst for Hempel, contain information that simply cannot be expressed propositionally (as in the case of diagrammatic representations of mechanisms).

As an alternative to Hempel's views on the nature of scientific knowledge, I propose to shift our philosophical attention to the epistemic activities through which models are handled and theories are created and interpreted. I emphasise the dual nature of biological knowledge as encompassing both theoretical knowledge ('knowing that') and embodied knowledge ('knowing how'). In the final section, I outline a working definition of the notion of scientific understanding that will be used throughout this thesis and hopefully starts to clarify the difference between theories and explanations (and other types of theoretical knowledge) and understanding. This is the idea of understanding as an ability possessed by individual scientists: that is, the ability to coordinate their theoretical knowledge with their embodied knowledge.

[14] For a succinct overview of Hempel's views on the matter, see the paragraph starting with the following sentence in 'Aspects of Scientific Explanation': 'in general, an explanation based on theoretical principles will both broaden and deepen our understanding of the empirical phenomena concerned' (1965, 345).

[15] One of my referees rightly encourages me to clarify that a practice-based approach does not represent the only alternative to the Hempelian framework, which has been critiqued from several angles (including for instance Bas van Fraassen's constructive empiricism [1980] and Peter Achinstein ordinary language philosophy [1983]). My focus on practice-based critiques stems from my present interest in the actual conditions under which biologists explain and understand phenomena.

[16] Hans Radder suggested to me that any science containing *reproducible* observations or experiments can be claimed to contain an empirical law in Hempel's sense. I do not agree with this objection, since many experiments in biology are reproducible not on the basis of their results (i.e. the correlations between phenomena for which they provide evidence) but on the basis of the activities involved in carrying them out (which cannot arguably be expressed in the form of law-like statements).

2.2.1 Theory-Pluralism

Can all biological knowledge, at least in principle, be subsumed under one set of law-like generalisations, as Hempel would have it? The reply to this question can only be negative, given the extreme pluralism characterising the perspectives, results, methods and aims to be found in contemporary biological research. Biology is growing increasingly disunified. Biological research is fragmented into countless *epistemic cultures*[17] (Knorr-Cetina 1999, 1), each with its own terminologies, research interests, practices, experimental instruments, measurement tools, styles of reasoning, journals and venues.[18] These cultures are typically constructed around a specific set of issues, the investigation of which requires training in a series of techniques, software applications and instruments that are regarded as appropriate to this aim.[19] Researchers enrolled in one of these cultures develop in-depth expertise about a specific set of topics and approaches, without necessarily complementing it with a broader awareness of which research is carried out by other epistemic cultures. It occasionally happens that a biologist shifts among cultures, thus acquiring overlapping expertises: still, this is a difficult and risky choice with regard to his or her future career, and allegiance to more than three or four cultures proves practically unattainable.[20] Because of divergence in epistemic cultures, each research community tends to develop increasing expertise on a very narrow topic. Dialogue among communities is difficult and sometimes nonexistent, as individual researchers have little interest, motivation or time to invest in communication with individuals with differing expertise. Further, as we shall see in more detail in the course of my analysis of integration efforts in the Arabidopsis community, extreme specialisation makes it hard to find common vocabularies, references and tools through which to understand each other. This means that many biological research groups happen to have overlapping research programmes, without however being aware of it and thus without gaining from each other's discoveries. Biology is indeed so fragmented a science, and so disinclined towards the construction of general theoretical narratives, that its disciplinary unity can better be traced via the objects that it studies (that is, living organisms) than via specific approaches or results (Cooper 1996, Keller 2002).

In this context, it is not surprising to find that different biological fields provide not only different theories about the same sets of phenomena, but actually different *types* of theoretical results, ranging from the mostly descriptive knowledge gathered by experimental or field biologists to the largely speculative research pursued by theoretical

---

[17] Epistemic cultures are defined as 'those amalgams of arrangements and mechanisms – bonded through affiliation, necessity, and historical coincidence – which, in a given field, make up *how we know what we know*' (Knorr-Cetina 1999,1). I will come back to this definition and its significance in practice in section 2.3.2.

[18] Careful analyses of each of these elements, aimed at emphasising methodological divergence and epistemic diversity among and within scientific disciplines, are proposed by Rosenberg (1994), Dupré (1993), Winther (2003), Bechtel (1993) and Mitchell (2003), among others. Hacking (1992), although not specifically focused on the life sciences, provides an enlightening view of styles of reasoning in science.

[19] For the mechanisms through which specific tools come to be regarded as indispensable within any one epistemic culture, see Knorr-Cetina (1999) and Galison (1999).

[20] More on the sociology of epistemic cultures can be found in section 2.3.

biologists through mathematical reasoning.[21] There is little chance that results obtained by different fields could be subsumed under a common framework, such as the one provided by Hempelian covering laws: this is not only due to the varying degrees of generalisability of results in each field, but also to the differences in the quality of results and in the explanations obtained (ranging from functional to mechanistic or causal). The field of population biology, for instance, strives to produce results in the form of mathematical definitions of the principles and laws guiding the evolution of life forms. Prima facie, these law-like mathematical statements seem to conform to Hempel's expectations of how a covering law in biology should work: they possess high generalisability and, potentially, great unifying power over the biological phenomena to which they refer. However, practicing biologists do not consider these results as covering laws, because their relation to the evolutionary patterns measured in actual biological populations is very unclear: data-sets and observation statements in more applied branches of evolutionary biology are formulated via reference to parameters that are incommensurable with the parameters used to mathematically describe general patterns in evolutionary dynamics. Similarly, there is little knowledge in physiology that matches the scope and potential applicability of an equation describing the behaviour of groups of organisms in population biology: what is achieved is an accurate narration of the mechanisms or processes giving rise to specific behaviours, structures or events. In the overwhelming majority of experimental life sciences, established explanations for biological phenomena contain much detailed information that is derived from data acquired through experiment and/or in the field. The law-like statements involved in such explanations tend to have low generality; in any case, they do not aim at unifying different phenomena, but rather at making sense of specific biological processes in relation to their local environment. Such detailed knowledge of the local processes is considered as an extremely important result in its own right, whether or not it is applicable to other cases and/or helpful to the construction of law-like generalisations. Given this context, the law-like statements produced in population biology cannot act as covering laws for the observation statements gathered by experimental biologists: there is currently no set of law-like statements under which the disunified approaches and results obtained in the life sciences can be subsumed and, thus, unified. In fact, as I shall illustrate in the next section, a whole apparatus of models, procedures and new results needs to be put in place in order to bridge the epistemological gap.

It might then be concluded, as Keller (2002) does in her own fashion, that the low epistemic value attributed to the unifying power of theories across biological disciplines is a defining characteristic of the biological sciences: extended fragmentation into different approaches and perspectives over the same phenomena does not allow to draw general analytic characterisations of what a theory is or should be in biology. The theoretical knowledge produced by the several epistemic cultures engaging in biological research takes a variety of forms, ranging from mechanistic explanations, to

[21] While wishing to emphasise the great diversity among the theoretical outcomes of different projects within each biological subfield, it is not my purpose here to engage in a survey of the notions of biological theory that have been formulated by philosophers of biology. An excellent critical survey is offered by Mitchell (2003). In Chapter 5, I discuss my own view on one type of theoretical framework used to develop integrative (plant) biology.

mathematical equations, to descriptions, representations and categorisations of phenomena. Researchers are not divided as to which type of theoretical formulation or approach is most informative or correct. More often than not, it is acknowledged that different types of theoretical knowledge can all have a *relative significance* towards the understanding of a given phenomenon (Beatty 1997, S433). As highlighted by Longino (2001) and Mitchell (2003), this pluralism in theoretical approaches and epistemic cultures is more conducive to the development of biological knowledge than a more uniform landscape, strongly centred on a few unifying covering laws, would be. Even the increasing concern with finding ways to integrate biological perspectives (which characterises research conducted in the Arabidopsis community, as documented in chapter 3) is both motivated by and appreciative of pluralism. Integrative approaches do not propose a unifying framework of law-like statements under which all types of theories can be subsumed (and which could, therefore, replace them). Rather, they are constructed as meta-frameworks to be used *in addition to* each culture's specific concerns, theoretical approaches and terminology, for the purpose of facilitating communication across different communities.

The *de facto* pluralism characterising both aims and results of biological research makes the Hempelian perspective on the nature of scientific knowledge look rather inadequate. Most biological fields, including molecular biology, genetics, ecology, developmental biology and physiology (to mention but a few), do not search for generalisations under which to subsume all data gathered through experimental or field activities (they cannot, in Suppe's words, be 'fruitfully axiomatised'; 1977, 66). What those fields do produce are mechanistic, structural, functional, causal, phylogenetic and/or historical information about many aspects of organic life. Such largely descriptive, yet conceptually structured information can hardly be dismissed as a reputable source of knowledge. This situation prompts the rejection of Hempel's views on the unifying role of theory in biology and its replacement with a more liberal approach highlighting the epistemological advantages of theory pluralism.

This refutation is clearly based on the observation of contemporary science rather than on a rational analysis of what science should be like. This is all very well for my purposes here, since I am interested in 'de facto' scientific knowledge (knowledge as it is actually produced and used), rather than in elaborating a vision about the 'essential' nature of knowledge or what scientific knowledge should be like (according to some rational and/or ideal principles). A devout Hempelian might justifiably raise a forceful objection against this attitude: Hempel's is a normative view and does not imply that theoretical knowledge as used in actual research always abide to the proposed ideal. Indeed, Hempel lists several ways in which his ideal is distorted in scientific practice (which is by definition incomplete, as it is in the process of extending and improving currently available knowledge), yet notes that the production of law-like statements remains the goal towards which scientists should and do strive (1965, 340). Several studies of biological practice found, however, that most biologists do not even strive to create unifying explanation of phenomena (and certainly not by reference to law-based generalisations): rather, they aim at understanding biological processes in their specific context, if possible by reference to various theoretical perspectives so as to understand several aspects of phenomena at the same time.[22] The pluralism characterising biological results and views on biological phenomena is happily tolerated, as long as each researcher is free to pursue her or his favourite approach and obtain results of interest to her or his peers and sponsors. As some have remarked, some parts of biology might even be regarded as 'science without laws'.[23]

2.2.2 Modeling Activities

As a step towards proposing an alternative view on biological knowledge, I now turn to the relation between theory and practice in contemporary biology. Rather than insisting on a specific definition of what a biological theory is or should be, I examine the ways in which theoretical knowledge is expressed, constructed, modified and interpreted by researchers. As it turns out, theoretical results are often expressed and handled by researchers in forms other than propositional statements: apart from symbolic representations (such as in mathematical equations), we find pictures, diagrams, photographs and various kinds of three-dimensional objects playing important roles not only in the context of discovery, but also towards the expression and justification of biological knowledge. Biologists refer to these representational tools as *models*.

Both biologists and philosophers acknowledge models to be crucial epistemological components of scientific reasoning and experimentation. Much has been written on the way in which models are produced and used to acquire theoretical knowledge about biological phenomena. Model-based reasoning, broadly defined as the use of representational tools towards gaining epistemic access to natural processes, is now widely recognised as playing a crucial role in the production of scientific knowledge. Further, it is acknowledged that models come in an endless variety of forms, a combination of which is always required by their use in scientific practice.[24] To investigate the epistemological role of any one model, we need to recognise and explore how and why biologists choose and handle specific combinations of them in order to pursue a given research goal.[25]

---

[22] For instance, see Winther's analysis of the difference in style and method between what he calls 'compositional' biology and 'formal' biology (2003, 2006). Rose (1997) illustrates how such theory pluralism works by means of a suggestive example. He lists the various ways in which a frog's jump into a lake can be explained: each of the five relevant explanations that he considers is compatible with the others, but cannot be replaced by them without significant loss in the understanding of the phenomenon in question.

[23] The recent volume edited by Creager et al (2006) under the title 'Science Without Laws' provides some excellent studies supporting this claim. See also Rosenberg (1994) on what he calls law-less biology.

[24] The collections of essays edited by Morgan and Morrison (1999), de Chardaverian and Hopwood (2004), Laubichler and Müller (2006) and Suarez (2006) provide excellent examples of the different types of models used by scientists and the combinations of different types required to acquire knowledge about natural phenomena.

[25] Much recent philosophical work addresses directly the piecemeal nature of modeling practices. For instance, Cartwright's framework is unique in its attempt to deal with different types of models across the natural and the social sciences (1999), while Suarez carefully considers the pluralism among notions of theory idealised within models (in Morgan and Morrison, 1999). Bailer-Jones published both on the different ideas that practicing scientists hold on what models are (2002) and on the use of 'sub-models' in

This position is well represented by the view recently proposed by Morgan, Morrison and their associates in their 1999 edited volume 'Models as Mediators'. Their analysis focuses closely on the practice of modeling and it does take interest in the overwhelming evidence for the diverse types of models used in scientific research. Making sense of the multiplicity of models and their uses is a major goal of the mediators view, which is one of the reasons for the broad definition given of models themselves: anything used by practising scientists to mediate between theory and phenomena can be called a model. The notion of mediation is used to suggest that a model serves 'both as a means to and as a source of knowledge' (1999, 35). I take this to mean that a model functions as *representative of* one or more phenomena as well as *representative for* a given theory taken to apply to such phenomena.[26] Models constitute the meeting point between knowledge and reality, thus providing 'the kind of information that allows us to intervene in the world' (1999, 23). By representation, Morgan and Morrison do not necessarily denote some kind of mirroring relation or structural (isomorphic) similarity between what is represented and the representation itself.[27] Rather, they imply 'a kind of rendering – a partial representation that either abstracts from, or translates into another form, the real nature of the system or theory, or one that is capable of embodying only a portion of a system' (1999, 27).

This account is very permissive, encompassing a potentially enormous diversity of types of representations. I see this breadth of scope as one of its main strengths.[28] Research in

---

order to obtain an overall model of a phenomenon (2000). A further example is represented by Levins' foundational work on biological models, which I shall consider closely later in the chapter.

[26] My use of 'representative of' and 'representative for' differs from the use proposed by Morgan (2003, 230). In Morgan's account, 'representative of' stands for the actual phenomenon that a model is taken to represent and 'representative for' refers to the range of phenomena to which the model can be further applied. In other words, Morgan's categorisation focuses on the range of typicality granted to the model. As I shall show in the case of Arabidopsis modeling, what matters in my account is not the extent to which a model is taken to be typical (as I shall illustrate, that depends strongly on the social and scientific context in which the model is used); rather, I focus on the degree to which the model of a phenomenon incorporates elements from a theory and thus makes it possible to apply that theory to the phenomenon in question.

[27] The characterisation of the relation between model and phenomena as isomorphic found a strong proponent in Suppes (1969, 2002), among others. For forceful criticisms of what Giere refers to as the 'instantial view' on models, see Giere (1999) and Suarez (1999).

[28] Notably, this very feature also makes this notion of representation rather slippery, for how are we to draw clear distinctions between a representation and 'the real nature of the system or theory' that it represents? Jeff Shank's analysis of modeling within behavioural biology (Shank and Koehler, 2006) constitutes a forceful illustration of this problem. In his view, most elements used by practicing biologists as tools for investigation – *including* theories, data and material samples - should be thought of as models, that is, as representations conveying some (partial and tentative) information about the phenomena that they represent. This argument comes from the recognition that, within any research context, the stipulation of what the phenomenon and theory in question are depends on the tools used to investigate them just as much as the choice and handling of those tools is based on what scientists endorse as theoretical hypotheses and descriptions of reality. How should we then draw boundaries between the notion of 'mediators' and the notions of theory and/or phenomena that they are supposed to mediate? The risk of a slippery slope, leading to the denial of the very possibility to distinguish between the model and what it models, represents a very serious problem for Morgan and Morrison. It runs counter their basic intuitions, which I share and value, that (1) an analysis of the representational value of models necessarily includes making explicit what it is that models represent at any given moment of the research process; and (2) models can represent theories as

biology involves the manipulation of several types of models, each of which is irreducible to all others and arguably plays a unique role towards the achievement of scientific understanding. We find discussions of *mathematical models*, *geometrical maps*, two-, three- or four-dimensional *computer simulations*, *robotic models* and *model organisms,* both simulated and real. Each of these types of model is constructed – or even 'discovered', as is arguably the case with model organisms[29] – in different ways and for different purposes. Their representational value (that is, what they are taken to be representative of/for) depends on their material features as well as on the interests and beliefs characterising the research context in which they are used. Further, none of those types of models is ever used on its own: all relevant biological literature underscores the need for a combination of different types of models in order to answer specific theoretical questions.[30]

This latter point is best explained by highlighting the constraints on the number of epistemic virtues that any one model can possess. Philosopher and biologist Richard Levins characterises those virtues as features that a biological model is expected to maximise in order to accomplish its mediating function in the most successful way. These features range from the requirement of generality (denoting the range of phenomena to which the model can be applied) to the idea of tractability (the ease with which it can be computationally or experimentally manipulated) and realism (the degree of empirical accuracy with which it represents relevant elements of the phenomenon under investigation). Levins' important observation is that no single model can posses all virtues at once. Emphasis on one of them is only possible at the expense of some other: for instance, all concessions to the realism of a model unavoidably detract from its tractability or generality. *Trade-offs* among the epistemological merits of models are therefore unavoidable: scientists need to employ a 'multi-modeling approach' involving the integration of several different models, each of which maximises at least some of

---

well as phenomena. Morgan and Morrison try to avoid the problem by a sophisticated discussion of what they mean by 'theory' and 'phenomenon', a detailed assessment of which eludes my present purposes (see Morgan and Morrison (1999) and Cartwright (1983, 1999)). In contrast, other scholars propose to embrace the problem by resorting to a 'modified' version of the semantic view on modeling, according to which scientific theories are in fact collections of models and their robust consequences, as coordinated by a theoretical perspective (Wimsatt 1987; Griesemer 2000). I do not agree with this latter solution. As I explain in the rest of this section, I view modeling as encompassing a set of activities which are not themselves theoretical, but which lead to the formulation of theories about phenomena (which, in turn, need to be interpreted in the light of the models used to create and apply them). Regrettably, providing a full argument for this crucial point would require writing another dissertation: I will therefore set it aside for now and hope to develop it in future work. I thank James Griesemer for discussions on this point.

[29] See Chapters 3.2.3 and 6 on this.

[30] Historians of science contributed extensive evidence for the importance of combining different models for the development of scientific research. For instance, some controversial cases of theory-choice characterising the history of physics and biology are being usefully reconstructed and understood as depending on changes in the range and interpretation of empirical evidence as much as on creative transformations in the number and sophistication of the modeling practices favoured within each of the rival theories (see for instance de Regt (2001), Keller (1999, 2000), Boumans (2001), Chang and Leonelli (2005a/b)). Equally indicative are the analyses of the evolution of research systems presented by Kohler (1991), Rheinberger (1997), Griesemer (2001), Galison (1997) among others.

those dimensions.[31] This allows biologists to optimise their overall modeling strategy at any given moment by constantly shifting and calibrating priorities from one set of requirements to another, depending on the changing demands of their research context (in turn dictated, at least in part, by the pursuit of the above-mentioned research outcomes). As Levins concludes: 'it is desirable, of course, to work with manageable models which maximise generality, realism and precision toward the overlapping but not identical goals of understanding, predicting and modifying nature. But this cannot be done' (1984, 19).

Given the necessity of a dramatic diversity among biological models, and the resulting 'promiscuity' of the notion of model itself in biology (Griesemer 1999, 436), much philosophical attention has been showered on the actual models employed in biological practice, so as to clarify the epistemological status of each type of model as both a product of and a tool towards scientific theorising. Relatively less attention has been devoted to the variety of activities that need to be performed in order to handle models adequately. Scientists do not only refer to models in their theories and explanations: they use them, manipulate them and modify them all the time in order to achieve and justify those very theories and explanations. The adequacy of such use is determined both by the features of the phenomena under scrutiny and by the material, social and institutional setting and commitments of the researchers involved. In a recent comment on this situation, James Griesemer suggested that 'knowing how models are made and deployed is as important to the philosophy of scientific representation as knowing that models are linguistically connected with phenomena by hypotheses in specified respects and degrees' (2004, 434). In other words, focusing primarily on modeling practices, rather than on models thus produced, might prove a useful way to gain insight in some long-standing debates in the philosophy of scientific modeling and representation.

Here I am therefore not interested in providing a general definition of what a scientific model is and how it functions in research practice (as Morgan and Morrison's position on these issues suffices for my present purposes).[32] My analysis relies on an account of

[31] This does not mean that there is one optimal combination of models to study each phenomenon, as the results expected of research through modeling vary depending on the interests, expertise and values of the group that carries out the research. For details on this view see Levins (1984), Puccia and Levins (1985), as well as the elaborations of Levins' argument presented by Wimsatt (1981, 1987), Griesemer (2000, 2006), Shank and Koehler (2006), van der Steen (1995, 26ff) and the relevant debate between Odenbaugh (2003) and Orzack (2005).

[32] Especially between the 1960s and the 1990s, a great part of the flourishing philosophical debate on scientific models concentrated on establishing which features, if any, are common to all models. This choice is understandable, given the need for at least a minimal definition of model that would allow to distinguish it from other elements involved in scientific experimentation and reasoning (and thus clarify its distinctive epistemological role). And indeed, it generated a vast body of literature fruitfully analysing the modalities by which models abstract, represent, illustrate and/or explain the aspects of reality and/or the scientific theories that they embody. Despite the interest displayed in the glaring differences among models and among the types of theoretical knowledge produced through their use, the importance, to practicing scientists, of using several types of models in combination did not receive much attention. The attempt to analyse the diversity among models turned into an effort to explain it away - either by arguing that only certain kinds of models should be characterised as such or by elaborating a unifying description of the processes of abstraction, representation and explanation characterising modeling activities. For instance, see Patrick Suppes' restriction of the definition of model to set-theoretical constructs (2002); Daniela Bailer-Jones' view of models as mental representations (2003); Hesse's early work on model-based

modeling that highlights the *epistemic activities* required to create, handle and interpret models.[33] Such analysis draws explicit connections between the different features of models and their role in securing one or more of the several epistemic goals of potential interest to practicing scientists. As illustrated in the previous section, that many different types of theoretical perspectives are used to study the same systems is certainly true of current biological sciences. Each way of theorising implies a different perspective, a different way of carving nature at its joints. Each type of research outcome arguably requires handling a different set of modeling practices. Hence the necessity to diversify modeling practices in order to comply with (or, as Levins would put it, trade among) the varying research outcomes and theoretical views characterising any specific research project.

Scientific modeling requires researchers to perform a series of actions in a *skilful* and *effective* way. By 'skilful', I mean that they have to resort to intellectual and material expertise in ways that are acceptable to their peers. For instance, epistemic activities required to construct and use models involve the selection and interpretation of the concepts and data to be represented in the models. Also, they include adequate handling of the instruments and/or software used to manipulate the model, together with the ability to create, interpret and/or follow guidelines and procedures set up to this aim. By 'effective', I indicate that the performance of those actions contributes to achieving the research goals of the community within which it is staged. Theories and explanations represent one kind of result of such performance.

It is not difficult to recognise a tension between my account and Hempel's exclusive focus on propositionally expressed knowledge. Not only models turn out to constitute an integral part of biological knowledge (as well as an important means towards its expression and communication): the very ways by which they are created and handled constitute parts of scientific knowledge, whose importance is entirely unacknowledged in the Hempelian framework. Further, even when looking at models as the results of these processes (rather than at modeling activities themselves), it is evident that non-propositional models cannot be reduced to or transformed into propositional ones without a substantial loss of information. Models such as graphs and three-dimensional objects contain different information from models such as descriptions and equations. In fact, the epistemological merits of models varies also depending on their material features: as I shall demonstrate in detail in chapter 6, a digital model of Arabidopsis biology yields different insights than a material model of it, meaning that these two types of models can be combined in order to obtain as comprehensive an understanding as possible of the same phenomenon. The re-evaluation of non-propositional knowledge that follows from

---

analogical reasoning as heuristic to the development of theories (1966); and Cartwright's (1983) account of how features of representationally powerful models are abstracted from phenomena and idealised from theories (which she modified precisely by focusing on the distinctions among models in her later work).

[33] I take the term 'epistemic activity' from Hasok Chang, who uses it to frame his rich reflection on the notion of intelligibility in science (unpublished). His framework as a whole constitutes a fascinating analysis of the epistemic value of intelligibility, as well as of the relation between scientific ontology and practice. Note that, while much of my present reflections are inspired by Chang's framework, I use the notion of epistemic activities in a different way, that is, to denote actual sets of action that are (mostly intentionally) carried out by researchers in order to reach a specific goal.

the study of modeling practices constitutes as strong a reason to abandon the Hempelian view on scientific knowledge as the dismissal of covering laws as essential components of theories proved to be in the previous section.

2.2.3 The Dual Nature of Biological Knowledge

I hitherto proposed a view of theories encompassing the diverse types of theoretical knowledge of phenomena obtained in different subfields in biology; and of modeling as the skilled performance, by a group of scientists, of a set of activities geared towards the achievement of a given set of research goals. I also sketched how this approach contrasts with the Hempelian view on scientific knowledge. Now I can finally outline the implications of this description of biological practice and its results for my original concern about the nature of biological knowledge.

The intuition that theories and explanations constitute but a small part of scientific knowledge was already put forward in the 1950s by the prominent philosophers Gilbert Ryle and Michael Polanyi. Both men presented systematic reflections on what I shall refer to as *embodied knowledge* – 'know-how' in Ryle's words (1949) and 'personal knowledge' in Polanyi's (1958). Their research suggests that scientific knowledge never involves simply a theoretical content, that is, that which is known (as expressed, for instance, by theories and explanations). It also involves the ability to apply such theoretical content to reality, to interpret it in order to intervene effectively in the world as self-aware human agents.

Polanyi sees such ability as an 'unspecifiable art', which can only be learnt through practice or by imitation[34], but which is rarely self-conscious, as it needs to be integrated with the performer's thoughts and actions without swaying his or her attention from his or her main goals (1958, 54-55). This is why Polanyi refers to this knowledge as 'tacit': it is an ensemble of abilities that inform scientists' performances (actions as well as reasoning), without being directly acknowledged by the individuals who possess them. Polanyi furthers this interpretation by distinguishing between two ways in which scientists can be self-aware: 'focal' awareness, which is tied to the intentional pursuit of 'knowing that'; and 'subsidiary' awareness, which concerns the scientists' perception of the movements, interactions with objects and other actions that are involved by their pursuit. In this view, focal and subsidiary awareness are mutually exclusive: one cannot have one without temporarily renouncing the other. Switching from 'knowing that' to 'knowing how' and back is, to Polanyi, a matter of switching between these two ways of being self-aware. In fact, his characterisation of know-how as 'subsidiary awareness' relegates this type of knowledge to a secondary position with respect to 'knowing that' (1958, 55). The acquisition of theoretical content constitutes, according to Polanyi, the central focus of scientists' attention. Tacit knowledge makes it possible for scientists to

---

[34] To Polanyi, the difference between learning by practice and learning by imitation corresponds to a difference in the tacit knowledge that is acquired: skills (in the former case) versus connoisseurship (in the latter case; Polanyi 1958, 57).

acquire such theoretical content, but it is not valuable in itself: it is a means to a (theoretical) end.

While drawing inspiration from Polanyi's emphasis on tacit expertise, I do not think that his hierarchical ordering of types of knowledge ('knowing how' being subsidiary to 'knowing that') does justice to the intertwining of 'knowing that' and 'knowing how' that characterises biological research.[35] Polanyi bases his reflections on the study of theoretical physics – and indeed, it might be that at least some theoretical physicists value theoretical knowledge much more highly than the practices and skills used in order to construct it. Within most of biology, however, the development of skills, procedures and tools appropriate to researching specific issues is valued as highly as – or, sometimes, even more highly than – the achievement of theories or data that it enables. The evolution of tools and procedures is crucial to theoretical developments and biologists openly acknowledge this to be the case.[36]

On this point, Ryle's view proves more accommodating. His analysis of the distinction between knowing that and knowing how does not assume either type of knowledge to be predominant over the other. This is due to an underlying difference between his outlook and Polanyi's: while the latter thinks of the distinction as descriptive, thus presenting the two types of knowledge as actually separate in practice, Ryle presents the distinction as a purely analytic tool. Indeed, he refers to the idea of an actual distinction between 'knowing how' and 'knowing that' as a false 'intellectualist legend': 'when we describe a performance as intelligent, this does not entail the double operation of considering and executing' (1949, 30). Also in contrast to Polanyi's analysis, Ryle defines 'knowing how' as involving both intentional actions and unconsciously acquired habits. In fact, he prefers an intelligent agent to avoid as much as possible the enacting of habits: 'it is of the essence of merely habitual practices that one performance is a replica of its predecessors. It is of the essence of intelligent practices that one performance is modified by its predecessors. The agent is still learning' (1949, 42). And further, 'to be intelligent is not merely to satisfy criteria, but to apply them; to regulate one's actions and not merely to be well regulated. A person's performance is described as careful or skilful, if in his operations he is ready to detect and correct lapses, to repeat and improve upon success, to profit from the examples of others and so forth. He applies criteria in performing critically, that is, in trying to get things right' (1949, 29).

Ryle does not specifically apply his account to the case of scientific knowledge. Nevertheless, his characterisation of agents 'trying to get things right' by iteratively improving their performance fits my characterisation of biologists' practices and interests, as reported in the previous sections. Biologists do indeed learn to use a variety of tools in order to pursue their intellectual interests, tools which I referred to as models.

---

[35] Note that, while Polanyi is widely credited for being the first philosopher of science to analyse the role of tacit knowledge in science, his classification of tacit knowledge as secondary in importance to theoretical knowledge has not received attention by his critics and followers alike.
[36] A nice illustration for this is provided by Edna Suarez's (2001) historical reconstruction of the role played by technologies (such as hybridization) towards the 'stabilisation of phenomena' in molecular biology. See also Rosenberg's views on the instrumental character of biological research (1994).

Biologists need to keep improving their ability to handle these models, in order to reach their goals. In the words of Mary Morgan, 'learning from models happens at two places: in building them and in using them. Learning from building involves finding out what will fit together and will work to represent certain aspects of the theory or the world or both. Modeling requires making certain choices, and it is in making these that the learning process lies' (1999, 386). As I shall discuss in detail in chapter 6, the function of models is not exhausted by their usefulness as thinking tools. These same models need to be handled adequately in order to incorporate and express theoretical knowledge about a phenomenon: in other words, they are tools for acting as well as for thinking. Thus, the knowledge acquired by a biologist successfully pursuing a research project encompasses both some results (like a theory, a set of data or a taxonomical system) and the procedures needed to develop, interpret and reproduce those results. A biologist is not born with the ability to perform the modeling activities required by his research. The skills and expertise enabling him to perform are important products of his work, in the same way as theoretical results are. In fact, as I shall contend at the end of this section, understanding those theoretical results becomes very difficult without reference to the know-how acquired via relevant research experience. Acknowledgements of the tight connection between learning and the material manipulation of objects have a long history and are increasingly accepted in current cognitive science.[37]

On the basis of these considerations, I characterise biological research as involving two types of knowledge, which are closely intertwined and virtually inseparable in practice, but which it is very useful to distinguish analytically. The first type includes what is typically characterised as the content of knowledge, that is what we regard as facts, theories, explanations, concepts concerning phenomena that are available independently of specific procedures or ways of acting. Access to cell theory and relevant data about cellular organisation, for instance, can be obtained without any expertise in cell biology or experimental research. I shall refer to this type of knowledge as *theoretical knowledge*. The second type is what I call *embodied knowledge*. This is the awareness of how to act and reason as required in order to pursue scientific research (for instance, by performing modeling practices). This awareness, whose features I shall characterise in more detail in chapters 5 and 6, is essential for biologists to intervene in the world, improve their control over the phenomena they study and handle the representations made of those phenomena. Embodied knowledge is expressed, for instance, through the procedures and protocols allowing scientists to intervene on the entities and processes of interest; the ability to implement those procedures and modify them according to the specific context of research; the acquired skill of handling instruments and models; the perception, often based on the scientists' experience in interacting within a specific space (such as a laboratory[38]), of how to move and position oneself with respect to the models and/or

[37] Indicative of this trend is the recent focus of fields such as artificial intelligence on the idea of 'embodied cognition' – see for instance Anderson (2003) and references therein.
[38] See also Latour's influential argument for any site of scientific research becoming a laboratory setting, in the sense of being standardised and perceived so as to eliminate all irrelevant elements to the research project(s) at hand (Latour 1988).

phenomena under study[39]; and the development of methods allowing to replicate experimental results.[40]

I thus posit that biological knowledge has a dual nature, encompassing both a theoretical and an embodied component. Before turning to the implications of this view for my account of understanding, let me note that this position nicely accounts for another widespread feature of contemporary biological research, that is the ever-increasing blurring of boundaries between 'applied' and 'pure' or 'theoretical' fields. I shall clarify this point via the example of biotechnology, to which I will come back when discussing the interests underlying Arabidopsis research (chapter 7). Biotechnology has traditionally been referred to as 'applied' research, because of its explicit focus on producing instruments, enhancing technological expertise and devising methods to apply theoretical knowledge to real processes (and thus produce a number of goods for use by non-scientists). A well-known instance of biotechnological research is the development of genetically modified organisms [GMOs], such as crops[41], which requires scientists to 'apply' knowledge coming from molecular biology and genetics in order to produce plants and animals with characteristics that are socially and/or economically advantageous. The classification of GMO research as applied does little justice, however, to the constant mix of theoretical and embodied knowledge required by biologists and bioengineers engaging in it, all of whom possess a high level of background knowledge as well as technical expertise. Technological expertise and specialised training are not only indispensable to the production of knowledge, but are part of biological knowledge itself. This is evident in the current distribution of research on the molecular mechanisms underlying bioengineering and the application of these mechanisms towards the production of useful mutants: most relevant laboratories engage in both of these areas at the same time. The large overlap between applied and blue-skies research explains why contract research has been steadily displacing governmental funding towards most biomedical and genomic research in the last three decades (Krimsky 2003). This is a blatant example of the impossibility to think of theories and explanations as disjoint from their intended scope and the way in which they are applied.

2.2.4 The Ability to Understand

I claimed that theoretical and embodied knowledge are two inseparable aspects of the knowledge used and produced by biological research practices. This view on the nature of biological knowledge is very different from the one offered by Hempel, who tried to formalise scientific knowledge by emphasising the axiomatic formalisation of its theoretical results (as Ryle and Polanyi would put it, its 'knowing that' dimension). I here finally turn to the implications of my critique to the Hempelian account of the nature of

[39] See Myers (forthcoming).

[40] Hans Radder (2002) convincingly argues that the development of methods to replicate experimental results is a necessary step towards the development of theories.

[41] Soybeans, maize, rapeseed and cotton are the only GMOs hitherto developed and marketed (mostly in the United States and India: Europe only admits maize up to now), due to current tight restrictions on the kinds of organisms that can be produced and sold in a genetically modified form (Freese and Schubert 2004; see the report published by Friends of the Earth [2006] for further bibliography on this issue).

scientific knowledge for Hempel's view on scientific understanding. As I remarked in my introduction to this section, Hempel classified understanding as a by-product of scientific explanations, thus implying that possessing explanations automatically leads to the acquisition of scientific understanding of the phenomena to which those explanations apply. In other words, Hempel claimed that access to theoretical knowledge leads to the understanding of that knowledge and of its applicability to real processes.

Now, it is certainly true that access to theoretical knowledge does not, per se, imply recourse to embodied knowledge. Scientific theories express the analytic concepts and categories employed in the study of phenomena through specific formulations, which can be propositional, symbolic (as in mathematical equations) or even pictorial (as in the depiction of a biological mechanism). These formulations can be independent of the use of models to develop and study them, as well as of knowledge about how to apply the theories. In this sense, theoretical knowledge can thus indeed be possessed without recourse to embodied knowledge. Using theoretical knowledge in order to understand natural phenomena is, however, an entirely different matter. Understanding phenomena by reference to scientific theories requires some level of know-how about how those theories apply to reality - such as the assumptions that are used, the models that have been employed to produce them, and so on. Hence, understanding natural phenomena by means of the theoretical knowledge accumulated by biologists does require some embodied knowledge (and vice versa, embodied knowledge needs to be coupled with theoretical interpretations in order to provide understanding): the individual in the process of acquiring understanding needs to intertwine his experience of the phenomenon (his observations and experimental interactions with it) with a theoretical interpretation of that same phenomenon. In view of this, I define understanding as *the cognitive achievement realisable by individual scientists depending on their ability to coordinate theoretical and embodied knowledge that apply to a specific phenomenon*.

By the verb 'coordinate', I refer to the variety of strategies that a scientist can learn to use in order to (1) select beliefs, thought processes and experiences that are *relevant* to the phenomenon in question and (2) *integrate* these components with the goal of applying them to the phenomenon.[42] I shall expand on some of these strategies in the course of this dissertation, with a special emphasis on modeling practices. This is not because I think that modeling is the only way to gather scientific understanding (in fact, as I shall specify in chapter 8, different types of understanding can be achieved, some of which derive from reflection, observation or taxonomic exercises, among other activities). Rather, it is because modeling is both a prominent activity in contemporary biology and a good exemplification of the means by which a coordination of knowledge that is relevant to a specific phenomenon can be achieved. In section 2.3.1, I illustrate this view on scientific understanding by means of two examples taken from contemporary biology.

[42] Galison uses the expression 'coordination of action and belief' to describe the interaction between experimentalist and theoretician culture in physics (1997, chapter 9). The practices to which he refers constitute, in my eyes, an excellent case for the analysis of understanding that I propose here: my definition of what is meant by coordination, as well as the notions of theoretical and embodied knowledge, are however largely unrelated to Galison's account.

2.3 Towards a Social Epistemology for Biological Understanding

> Of course, the interactions between the body and the environment are not the only form of cognitive mediation with the world. It is obvious that social communication too (and of course manipulations of the conditions of possibility of this communication) provides rich information to many of our senses.
>
> Lorenzo Magnani 2001, 68

Once we reconsider the role of the individuals conducting research, thus doing away with Hempel's rejection of what Popper called 'psychologism', we are immediately drawn to asking questions about the characteristics of such individuals, the way in which they share experience and goals among each other and participate in collective projects. Especially in the context of 'big science' projects, such as Arabidopsis research today, the constant exchange of insights, data and resources among researchers is not an option, but a necessity. Even more importantly, the organisation of scientific research into distinct communities, and its significance to scientific epistemology, is as widespread a phenomenon as it is understudied. What are the epistemological implications of carrying out research in large, multicultural and multi-sited groups? What is the relation between individual scientists and the communities in which they work? Does it make sense to think of scientific epistemology as a social phenomenon? And what types of socialisation does research in communities actually involve?

This section examines some of these questions, as relevant to my current investigation of scientific understanding. While the characterisation of understanding offered in the previous section emphasises the role of individual researchers in understanding biological phenomena, I certainly do not intend to underestimate the extent to which individual understanding is shaped by her or his membership in a scientific community. As I am about to clarify in this section, I take understanding to be 'scientific' precisely when it arises from an agreement, reached among the relevant actors, over what constitutes adequate theoretical and embodied knowledge. Such an agreement need not be explicit, as many instruments and protocols are adopted and recognised as valid without formal approval by the community. Also, the resources, tools and concepts in use in a scientific community do not need to cohere with each other in any way, since in the same community very diverse concepts, techniques and instruments might be used: each scientist is free to associate tools and resources as wished, depending on his or her preferences and beliefs.

Given these caveats, I wish to argue that scientific understanding is a socially nurtured phenomenon. Moreover, the notion of 'social' applying to the case of scientific communities is not just the minimal characterisation involving the interaction of at least two individuals (thus, 'social' as 'interpersonal'). In the vast majority of contemporary scientific communities, social norms, behaviour and forms of communications are dictated and enforced by institutions or other super-individual organs. This implies that for scientific understanding to be social, it needs to be sanctioned by institutionalised measures such as peer-reviewed publications, public presentations and receipt of research grants: the understanding of a phenomenon shared by two or three individuals only becomes scientific when it is expressed to other members of their community through means independent of any one individual and instituted in order to control the quality of the research in the community as a whole.[43]

Prima facie, the study of scientific understanding as an essentially social phenomenon seems almost to contradict the definition of understanding given in the previous section. When I characterise understanding as the ability to act and think in a specific way, I seem to make this ability depend entirely on the individual who is acting and thinking in such a way. Yet, this is not the case. This section illustrates that the individual's ability to understand, as well as the quality of that understanding, is largely shaped by the participation and allegiance of that individual to the epistemic culture of one or more research communities. This argument is exemplified in detail in chapter 7, which is devoted to an analysis of how the structure and power relations characterising the Arabidopsis community influence the scientific understanding that its participants have of Arabidopsis biology.

2.3.1 Individual and Communicative Understanding

> *The account of cognitive agency as involving interdependence means that individuals know to the extent they interact critically with others in cognitive communities. The context-bound character of validation and the characterisation of cognitive agents as interdependent do not deny the importance of individual agents in the construction of knowledge. It does mean that attributing knowledge to them is attributing to them some relation to their cognitive communities, as well as to the objects and content of knowledge*
>
> *Longino 2002, 122*

Several philosophers of science emphasise individual work, responsibilities and discoveries when accounting for the history and structure of science. More or less explicitly, they maintain that scientific epistemology is not inherently social and that the study of science requires focusing on individual achievements rather than on the environment that frames these individuals' actions, intuitions and decisions. This position informs many accounts of scientific discovery, where few individuals are held responsible for the birth and establishment of new ideas; it also permeates most analyses of scientific reasoning and methods, which are, after all, enacted by individuals. Here I would like to counter this approach by suggesting that there are epistemologically relevant differences between research conducted by an isolated individual, the research

[43] In this broad characterisation of what I mean by 'social', I take inspiration from Durkheim's intuition about what makes a fact 'social' at all: that is, 'ways of acting, thinking, and feeling, that are external to the individual and endowed with a power of coercion, by reason of which they control him' (Durkheim [1895] 2003, 27).

conducted by an individual working in a small group and the research conducted by an individual working in a large group whose work is regulated by institutionalised norms.[44] Most contemporary research, certainly in biology, is indeed pursued in large communities. The organisation of a scientific community can be examined according to a multitude of variables, each of which commends a different partitioning of the community in question. These variables include disciplinary expertise, common sets of skills, degree of intellectual authority and/or credibility, level of education and power status within and without the community. How does this complex social organisation influence the individual's ability to understand? And where can we start in order to tackle this issue?

Let me start from a common-sense interpretation of the notion of understanding, which certainly underscores its individualistic nature. This is the idea of understanding as a sudden insight, an intuition exemplified by what Keller calls the '*Aha!* Feeling' of cognitive grasp (2002, 12). In his 'Philosophical Investigations', Wittgenstein reflects on the notion of understanding implied by utterances like: 'Now I understand!'. Passages 150 and 151 report a distinction between such a notion of understanding, which is tightly connected to individual imagination and creativity, and a different notion of understanding, which is used in interpersonal communication to denote the ability to do something. Wittgenstein does not elaborate on this intuition, yet there is much to be drawn from it. What I intend to argue here is that understanding of the 'Aha!' type is not, per se, scientific. This is where Hempel and his followers had it right. This type of individual understanding consists in the acquisition of an intuition that is triggered by the individual's personal experience and leads him or her to critically reconsider the knowledge that he or she already possesses: this intuition is a subjective interpretation of reality, which is typically derived by the individual's interaction with phenomena or their representations. Examples for this type of understanding are typical cases of discovery both in everyday life (such as pushing a door and understanding that it is open, or trying an unknown road to work and understanding that it is actually a shortcut) and in scientific research (doing an experiment with animals and discovering that they behave in different ways from what was expected).[45]

[44] William Bechtel captures one of these differences when describing the transition of cell biology from a research field to an actual discipline (that is, an independent unit with its own locations, venues, journals and internal dynamics): 'in order to make cell biology a viable discipline, and not just a research domain to which a variety of disciplines might contribute, it was necessary not only that appropriate research tools be developed and that new information be derived from use of the tools, but also that institutions emerge which could provide a stable work environment in cell biology' (1993, 292).

[45] I do not mean to treat scientific understanding as synonymous with scientific discovery, or at least not with the traditional definition of discovery as an event that brings new knowledge to a given community (in the form of a new object, concept, theory, result and so forth). What I do maintain is that an individual's acquisition of understanding can be seen as a form of discovery in a different sense: that is, as an event that brings new knowledge to that individual, even when it does not add anything to the knowledge possessed by the community of which that that individual is a member. Whether that individual's new understanding will be counted as a discovery by the whole community depends on whether the community accepts that understanding as (1) truthful and (2) innovative.

The 'Aha!' type of understanding can be sought for, but it is not achieved intentionally.[46] It is not possible to predict when that intuition will be experienced, under which circumstances and with which implications. The elements of unpredictability and randomness characterising individual understanding highlight the creative, non-intentional nature of this cognitive process. They are also responsible for the fascinating mix of progressiveness and mystery (even, sometimes, serendipity) that characterises scientific pursuits. For these same reasons, individual understanding is a necessary but insufficient condition for the achievement of scientific understanding. The latter emerges solely from each individual's attempts to share his or her intuitions with other researchers and explore the significance of that finding in the light of knowledge previously acquired by a group of individuals devoted to study similar topics.

Why should this be the case? Consider the example of scientific discovery given above, in which an experimental biologist experiences the 'Aha! feeling' of understanding when witnessing an unexpected behaviour by laboratory animals. The renowned developmental biologist John Bonner reports exactly this sort of experience in his memoirs. He recounts how an unexpected behaviour by his favourite organisms, the slime molds, triggered his intuition about chemotaxis[47] and effectively started his proficient career as a scientist:

> What surprised me [after the discovery] was how quickly I read the message sent to me through that one glance into the dissecting microscope. *Instantly I saw that chemotaxis had been proved,* and that I made the discovery that would get me a satisfactory thesis. I remember dancing about my lab room and punching the air in my excitement. The experience also taught me a great lesson. I had not carefully designed an experiment that would prove diffusion; I had managed it by accident. That and all the other observations I had made told me that the slime molds were in charge, not I. They would let me know their secrets on their terms, not mine. A gifted and delightfully eccentric mathematician who helped me with the publication of these results knew the same thing: he would write an equation, stare at it for a bit, and then as though I were not in the room, he would say to the equation, "speak to me, speak to me". Well, the slime molds had spoken to me (Bonner 2002, 77-78; my emphasis).

Bonner's 'Aha!' experience seems to be entirely individualistic and dependent on his own interaction with his objects of study, the slime molds – in exactly the same way as opening a door or discovering a convenient shortcut to work is, as in my previous examples. How can Bonner's experience of understanding be defined as social? In order to answer this question, let us consider what it is that Bonner actually understands here. What Bonner understands is not the pattern of behaviour of the slime molds per se, but rather its relevance in the context of existing biological research on chemotaxis. Before his enlightenment, Bonner already had *theoretical knowledge* about what the biological

[46] See Kleiner's debate with Nickles over the degree of premeditation and intentionality involved in scientific 'discovery' (1999).
[47] By chemotaxis, biologists indicate the process through which single-cell or multi-cellular organisms direct their movements in response to the presence of specific chemicals in their environment (allowing, for instance, bacteria to find food or flee poison).

community meant by chemotaxis. He was well trained in what more senior biologists thought of the phenomenon and of its implications - otherwise he would not have been able to 'instantly' recognise it in the behaviour of the slime molds. Through his weeks of laboratory experiences with slime molds, he had also become familiar with their habits, behaviours and reactions to various experimental set-ups, thus effectively accumulating some *embodied knowledge* about how chemotaxis might manifest itself. Thus, slime molds might have unexpectedly 'spoken' to Bonner: but what allowed him to 'listen' to them were his combined abilities to phrase their message in a specific terminology, match it to relevant laboratory experiences and interpret it according to the theoretical background that he had been taught by his supervisors. Bonner's understanding of how chemotaxis occurs was enabled by his participation in a research community as much as by his own ability to use the tools acquired in that community.

As I pointed out in the previous section, understanding results from the ability to coordinate theoretical knowledge with embodied knowledge. In the case of Bonner's discovery, this means that his new observations on slime molds behaviour were assessed through the spectacle of the theoretical knowledge that he already possessed (such as previously accumulated speculation about chemotaxis) as well as his embodied knowledge about the usual behaviour of slime mold, as he experienced it day-in, day-out in the laboratory. Bonner's experience of scientific understanding is made 'scientific' by its continuity and intertwinement with the work conducted by the whole community of biologists working on the same type of phenomenon. This is a first sense in which understanding needs to be social in order to be scientific: both the theoretical and the embodied knowledge used in order to make sense of the new observations are the fruit not of Bonner's personal intuition and experience, but of the agreement reached among several individuals about which terms, models, issues and instruments to use in order to examine the phenomenon that is being understood. Notably, not any individual or community will do: the agreement among individuals needs to be sanctioned in at least one scientific community, in the sense of being incorporated in the series of methods, instruments, concepts used by members of that community to express their knowledge and pursue their research. This last point, to which I shall return in the second half of this section, is important since many types of everyday understanding can be claimed to depend on social constraints of some sort (think of the awareness of social codes and use of language characterising a specific community that is needed to understand a joke). What makes understanding scientific is the extent to which it is nurtured by knowledge accepted and cultivated in one or more scientific communities. An isolated individual might have the deepest and most meaningful understanding of a specific biological phenomenon. However, that understanding can hardly be categorised as scientific unless it is informed by current scientific knowledge of that phenomenon *and* unless the individual in question attempts to offer that result as a contribution to one or more established scientific fields.

This last statement already contains a reference to another sense in which I think that scientific understanding is social. This second aspect concerns the necessity for a community, in order to build such consensus over what constitutes acceptable theoretical and embodied knowledge about the phenomena under scrutiny, to communicate and assess individual results at the community level. Understanding becomes scientific when it can be communicated, either by speech or through the imitation of others' actions. In order for an individual's understanding to be relevant in a research community, she needs to refer to a linguistic formulation, a series of representations and/or specific experimental set-ups that will enable her to communicate her understanding to her peers.[48] In contrast to everyday understanding, scientific understanding captures the process of finding ways to exchange knowledge about nature among human agents.

In the case of Bonner's discovery, this means that his understanding would not have a scientific status unless he disposed of the right tools, terminology and skills to allow other biologists to share and replicate his experiences. This second aspect of the social nature of scientific understanding is clearly illustrated by reference to a very recent case of biological controversy, whose details have peppered the pages of journals and newspapers all over the world. This is the debate on the credibility of the results published by Hwang Woo-suk, an internationally recognised expert in stem cell research and the head of a prestigious laboratory at Seoul National University, South Korea. At the beginning of 2004, Hwang Woo-suk and his American associates announced that they were about to publish a paper in 'Science' containing findings on human somatic cell cloning that had the potential to revolutionise stem cell research (Hwang, 2004). Immediately after the paper was published, several laboratories tried to replicate Hwang Woo-suk's findings, without success. A huge controversy erupted on the techniques and procedures used by the South-Korean lab in order to achieve its results. The controversy started from questions, raised by one of Hwang Woo-suk's colleagues in 2005, about the ethical legitimacy of the manner by which the necessary eggs had been acquired from female donors. Doubts then spread to the procedures with which the experiments had been carried out, as well as on the quality of the documentation provided. It was claimed that the procedures were impossible to reproduce; that data had been falsified; and that the photographs of the process provided by the South-Korean team had been digitally altered to fit their intended findings. In other words, the international community of stem cell researchers that had so enthusiastically bought into Hwang Woo-suk's results started to question the validity of his procedures: the embodied knowledge displayed when producing those acclaimed results was found to be inadequate to the standards imposed by the community on its individual members. By the end of 2005, the clash between Hwang Woo-suk's procedures and the ones recognised by the international community proved so powerful as to totally discredit Hwang Woo-suk's results. He was forced to step down from his job and issue a public apology. Even more interestingly in the light of my arguments, the editors of 'Science' apologised for having published his results (Chong, 2006), an action that was widely taken to exemplify the unreliability of peer-review processes conducted by major scientific journals (Bosman, 2006). The efforts and personal humiliation suffered by these editors in order to salvage the peer-review system is understandable, as this system constitutes one of the foremost institutional measures for a community to gauge the quality and adequacy of an individual's or a group's research *in the light of accepted standards in the field*.

---

[48] The modalities under which scientific understanding, as I define it, can be communicated will be explained in detail in chapter 7, section 7.2.1.

In its quality of 'ability to do something', understanding is a prerogative of individual experience. However, as illustrated by the above examples, the knowledge required to exercise this ability, as well as the standards placed on what constitutes relevant and adequate knowledge with respect to the phenomenon that is being understood, is a social achievement by whole research communities. The individual experience of understanding thus qualifies as 'scientific', only if it deploys (and eventually adds to) the theoretical and embodied knowledge that is already shared by the research community within which the individual's work is conducted and evaluated. This explains why scientific discoveries, when breaking loose from traditions already established, are so difficult to communicate to other scientists: the communication of new findings implies some break from the theoretical and/or embodied knowledge hitherto accepted in the relevant communities.[49]

When viewed from this perspective, scientific understanding could be defined as the cognitive substratum to what Thomas Nickles calls 'reconstruction', i.e. something that scientists 'must do [..] in order to apply old results and techniques to new problems at the frontier and to model one problem solution on another' (Nickles 1988, 34).[50] This account resonates with Thomas Kuhn's notion of paradigm, especially in the reading that Imre Lakatos gave of it through the notion of 'research programmes'. According to Lakatos, such programmes are constantly modified through ongoing research, with the exception of their 'core', which is composed of key concepts, beliefs and ways of acting that unify the work of any research community and make it possible for it to be progressive (Lakatos, 1970). James Griesemer rephrased this intuition by emphasising the importance of the knowledge that biologists take for granted while conducting their research. Griesemer refers to the core beliefs of a research community in biology as its 'theoretical perspective': that is, the set of concepts, interests and values that are (largely uncritically) used by biologists in their research and that, therefore, demarcate their epistemic culture. A theoretical perspective does not apply directly to a specific set of phenomena. Rather, it contributes the analytic and practical tools needed by a scientific community to pursue and obtain knowledge about a specific phenomenon (Griesemer 2000, S348; more on this in chapter 5, section 3).

It is precisely the constant inter-personal (and inter-group) communication of results and methods, as well as the institutionalised practices used in order to allow for such exchanges, that qualifies understanding as 'scientific'. In this sense, my approach to scientific understanding embraces the suggestion, made recently by Michael Friedman, that scientific research is tied to a *communicative rationality*, that is, to 'our capacity to engage in argumentative deliberation or reasoning with one another aimed at bringing about an agreement or consensus of opinion' (Friedman 2001, 54). The social roots of deliberation, which several scholars of science view as the foundation for scientific and

[49] Note that my account, despite admittedly Kuhnian in its emphasis on expertise, in no way implies the incommensurability of results, methods and practices used within different epistemic cultures. As I emphasise in the rest of my thesis, each individual scientist might participate in several epistemic communities at once; as a result of this, as well as of the constant evolution of research, the boundaries separating different communities are often permeable and fluid.

[50] Gooding phrases the issue as a question of scepticism toward one's own findings: 'This is non-Cartesian scepticism – a state of uncertainty that allows for reconstruction and the reinterpretation of one's own experience so as to make it compatible with aspects of experience as reported by another' (1990, 21).

technological knowledge, are what Helen Longino refers to as ‘the rationality of social cognition’.

2.3.2 Collaboration in Scientific Communities: Focusing on ‘Big Science’

> *‘Collaboration’ as a term is helpful insofar as it indicates different individuals or groups aiming at certain shared goals, but we can and have gone further toward a specification of how the coordination takes place. Indeed, far from melting into a homogeneous entity, the different groups often maintain their distinctness, whether they are electrical engineers and mechanical engineers, or theorists and engineers, or theorists and experimentalists. The point is that these distinct groups, with their different approaches to instruments and their characteristic forms of argumentation, can nonetheless coordinate their approaches around specific practices*
>
> *Peter Galison 1997, 806*

After having clarified the reasons why scientific understanding can be regarded as a fundamentally social phenomenon, let us have a closer look at the organisation of research communities and the implications of different modes of scientific collaboration for the study of scientific understanding.

Knorr-Cetina’s work on epistemic cultures, to which I referred in section 2.2.1, departs from the observation that individual researchers depend on others in order to carry out their research. This might have not been the case (not to this extent, anyhow) in the 18th and early 19th centuries, when the study of science was still the leisurely occupation of rich gentlemen, rather than a profession in its own right. The professionalisation, and consequent institutionalisation, of scientific research meant that contemporary scientists could not afford isolation, whether physical or intellectual. Most of their work takes place in facilities whose use is shared with other researchers; their training is carried out within universities and requires extensive collaboration with both teachers and other students; their research is based on and connected to several other endeavours by senior scientists or colleagues; and, most importantly, the quality of their work is assessed with relevance to its role in and relation to related research carried out by others. The goals, methods and results of each scientist’s research depend heavily on the goals, methods and results of his or her peers; this is what is meant by *epistemic dependence* in science (Hardwig 1985).

Epistemic dependence among individuals belonging to research communities of various sizes may vary both in quality and in quantity. By quantitative variation of epistemic dependence, I mean variation in the extent to which an individual or research group depends on the work of colleagues. Similarity of interests and geographical proximity, for instance, might determine closer collaboration (even if this latter factor is losing importance in the context of globalised transport and communications, as I note below). Also, the degree of dependence of individuals on material and institutional constraints –

such as for instance shared access to a particularly expensive experimental apparatus, use of a common pool of funding or need to cooperate to obtain a grant – determines the quantity of epistemic dependence among individuals that are subject to them.

The quality of collaborations among scientists within a community can be classified according to the background and role of those individuals. Thagard proposes to distinguish among four types of collaborations on the basis of the individuals' institutional role and expertise, as follows:

(1) employer/employee
(2) teacher/apprentice
(3) peer-similar (that is, collaboration among scientists with similar expertise)
(4) peer-different (that is, collaboration among scientists with differing expertises).

This classification exposes the different characteristics of each type of collaboration (Thagard 1997, 245-6). For instance, both (1) and (2) usually involve strong asymmetries in knowledge, status and power characterising the agents involved. At the same time, (1) constitutes a weaker form of collaboration than (2), as an employer trains employees to execute his or her own orders (as in the case of laboratory technicians), while a teacher trains apprentices to think and act for themselves (as in the case of a postgraduate student). The asymmetry characterising collaboration (4), also referred to as 'interdisciplinary', is of a very different type from the one expressed in (1) and (2). It indicates that the specialised skills, training and knowledge of the individuals involved do not overlap (or at least, not to a great extent). At the same time, this lack of reciprocal understanding puts them on an equal stand: they both have to learn from the other in order to construct a framework in which they can share their expertise. This effort is not nearly as marked in the case of collaboration (3). This last type of collaboration might be the one involving the most epistemic dependence, as sharing of information and materials among scientists working on the same issues in the same way is often vital to the success of the relevant field. At the same time, this type of collaboration is the most vulnerable to competition for the limited number of powerful positions available within the field. Often, the advantages of competition trump the advantages of collaboration, thus leading to a secretive environment where results are rarely shared and collaboration occurs only when enforced by institutions.

Of course, Thagard's system is not the only way in which collaboration among scientists can be analysed. Other methods include noting distinctions between individuals working in 'central' institutions and individuals working in 'peripheral' institutions (Evans, 2004); between theorists and experimentalists, or field biologists and laboratory biologists (Galison 1987); and between scientists allied with specific (and sometimes opposing) scientific, social, economic or political interests (Fuller 2002). Also, Thagard's classification concerns mostly overt collaboration, while much scientific collaboration involves indirect contact with others. For instance, the peer-review system, as well as the referencing system adopted in scientific journals, does not presuppose (and sometimes precludes, as in the case of referees) that contributors know each other personally. An individual or team may collaborate in a large project by perfecting a protocol, model or tool, thus spreading expertise without directly interacting with who benefits from it. Indirect collaboration can be just as effective as overt collaboration among scientists, even if skills and tacit abilities (what I characterised as embodied knowledge) are most effectively acquired by close contact with individuals who are already experienced in the type of research in question.

The importance of indirect collaboration and collaboration at a distance is especially felt in relation to so-called 'big science' projects, to which I will devote the bulk of my analysis. Scientific research, especially in biology, is increasingly financed and structured around large projects involving overt collaboration and sharing of resources among various institutions. They are typically interdisciplinary, thus including various 'peer-different' collaborations. Given the specificity of the topics at hand, they include researchers based at different locations, often widely distant from each other (as in the case of American-Japanese collaborations, for instance). Steve Fuller gives a negative assessment of the effectiveness and organisation of big science projects. He remarks that, while a critical mass of contributors and resources is certainly necessary to conducting scientific research in any given field, contemporary large-scale science has little to do with constructing this critical mass and a lot to do with maximising the power, prestige and social influence of the individuals and institutions involved (2000, 33). According to Fuller, the complexity and hierarchy inherent to big science, not to mention the several social, political and economic interests bound to such expensive undertaking, represent a threat to the integrity and credibility of the science that is produced: 'many of the essential virtues in science, especially those associated with criticism and openness, seem to get lost once scientific institutions reach a certain size, complexity, hierarchy and level of material investment' (2002, 28).

Fuller's assessment touches a core concern in large-scale research regulation, which is often in the position of offering what Jasanoff (1990) has called 'serviceable truths' for policymaking (whether it is privately or publicly financed). Yet, I think that the type of collaboration fostered by big science has not only negative but also positive connotations. The reason why large projects are an increasingly popular way of organising all kinds of scientific research are not purely financial and political, even if opportunities for economic advantages and social prestige (especially in media resonance) abound in this mode of research. Another reason for the success of big science is its corrective influence over the lack of collaboration plaguing competitive fields like genetics and molecular biology, where both peer-similar and peer-different collaboration is avoided by individual scientists in order to enhance each individual's chances of publishing and climbing the career ladder (with negative consequences for dialogue and overall productivity of the field). In this context, big science projects prove effective in order to encourage groups at different locations to collaborate, coordinate their research strategies and share data and instrumentation.

The positive dimensions of big science are recognised by funding bodies, particularly at the governmental level. Different types of sponsorship give rise to mainly two modes of large-scale collaboration, which I shall call centralised and decentralised big science, respectively. *Centralised big science* is launched and coordinated by few (often only one) leading institutions, which acquire funding and distribute it among interested groups from other institutions on condition of complying with a given research agenda. As I shall illustrate in chapter 7, the Arabidopsis community constitutes a good instance of this mode of collaboration. The research conducted in the community is highly centralised, despite its geographical dispersion. The leading institution proposes a research program for all participants to follow according to specific conditions and rules, including the methods, tools and materials to be used. In turn, all participants report their findings to the leading institution, which thus acts as the scientific, financial and administrative 'centre' of the community. As is evident even from this short description, centralisation often entails a high amount of standardisation, which guarantees that all participant groups indeed follow the rules and conditions dictated by their sponsors and/or the leading institution.

In the case of a collaborative project among equally powerful institutions, research does not have to be centralised. Rather, it is often the case that all participant laboratories agree on the set of issues to be investigated, and then each laboratory carries out its own research without consulting with others unless it needs to share information or resources (the European Union often sponsors collaborations of this kind among laboratories based in different European countries). Results are then reported in international venues such as conferences and workshops, where participant researchers correct each other's inferences and try to integrate them into an overall understanding of the set of issues on which the overall project is focused. In order to start this type of collaboration, standardisation of practices and theoretical frameworks is not necessary, as each group pursues the common topic in its own way. In the absence of a priori constraints on the instruments and methods to be used, the chances for some of the groups to yield interesting and innovative results are higher. On the other hand, the lack of standardisation makes it more difficult for groups to collaborate and it generates problems when the time comes to integrate results by different groups. I shall refer to this type of community as an instance of *decentralised big science*.

By highlighting some positive features characterising centralised and decentralised big science, I do not mean to undermine Fuller's argument about the negative aspects of research carried out under such conditions. Indeed, I agree with him that in centralised big science it is more difficult to maintain a pluralism of perspectives: the high degree of standardisation, inter-group communication and the need to publish results jointly are only some of the characteristics of big science that threaten to homogenise and flatten the controversies that might emerge between different groups involved in a given project. As I signalled in section 2.2, there are reasons why so many biologists and philosophers deem theoretical and model pluralism to be an advantage, rather than a detriment, to scientific research: as forcefully argued by Peter Galison, among others, 'it is precisely the disunification of science that brings strength and stability' (Galison 1997, 781). As I shall discuss with reference to big science collaboration on Arabidopsis research, there are strong doubts among biologists as to whether it is at all possible to 'reconcile' protocols, results and techniques used by groups with differing perspectives, beliefs and interests to different aspects of the same phenomena. Further, the provisional agreement achieved in large constellations of powerful groups tends to have repercussions on the work of smaller communities, which are not part of the big science circuit but have to publish in the same journals, consult the same reviewers and appeal to the same sponsors.

In the long term, this input might be damaging to the creative pluralism that is inherent to biological research in precisely the way envisaged by Fuller. The construction of consensus especially within centralised big science can be extremely undemocratic, as more powerful institutions tend to simply impose their beliefs and methods to financially and/or socially weaker groups.

Even in the face of all these misgivings, however, I think that power struggles in big science project can be confronted and regulated in different ways, not all of which are bound to have negative effects for the development of science. Under some (especially centralised) arrangements, most useful controversies and debates will indeed be eliminated; under other arrangements, only the irrelevant and fruitless discussions will be cut out of scientific attention. Big science does not *necessarily* lead to the elimination of useful controversies on the results obtained by different groups involved in such research. Extensive collaboration among different epistemic communities also yields more opportunities than otherwise to exchange and debate perspectives and opinions: if dissent and pluralism among scientists are at all valued in a big science project, research carried out in such a context might actually foster, rather than undermine, the critiques and confrontations that proved crucial to scientific development throughout the history of science. Both centralised and decentralised big science foster collaboration over dissent among participants. Agreement must be reached on which topics to investigate and, especially in the former case, on which methods and instruments to use: as remarked by Star and Griesemer, ‘each protocol is a record of a process of reconciliation’ (1989, 407). Such reconciliation is helpful especially when the goal of biologists involved in big science research is, as in the case of Arabidopsis biologists, to attempt to integrate the all-too-many approaches and studies hitherto carried out on the same phenomenon (plant biology) into a unique framework. Many biologists view integrative biology as a step towards a better understanding of the sea of data hitherto gathered on all aspects of biological phenomena: we can really understand how an organism or an ecosystem functions, it is argued, only when we find ways to analyse many of its characteristics at the same time and under a common framework.

Integrative biology is now far from being achieved and attempts towards obtaining large frameworks of this kind can only be carried out in the context of big and interdisciplinary research projects. In this sense, the collaboration enforced in big science projects might be viewed as enabling biologists to exchange and debate their views in more efficient ways than a more local context where collaboration is totally lacking. Arguably, theoretical pluralism is today so extensively disseminated as to threaten the very possibility for biologists to examine the foundations and general framework for their research. Occasional reflection on what scientists agree on, rather than disagree upon, is relevant to acquiring a balanced vision on the state of the field. In the face of the unstable, dynamic, complex organisation of scientific research into networks competing for the same, limited resources, large-scale science thus provides a platform for collaboration and for building bridges between the results, beliefs and methods upheld by different groups. It provides a space for different communities to meet and build consensus over what makes for acceptable (theoretical and/or material) knowledge: it is thus, arguably, an important platform through which to enhance biologists' understanding specifically of complex biological phenomena.

My account of the tension between epistemological advantages and disadvantages brought about by 'big science' collaborations highlights the extent to which the research context (and specifically the degree and quality of epistemic dependence among individual biologists) influences both the theoretical and embodied knowledge used by researchers to understand biological phenomena. The quality of their understanding, as provided by coordinating theoretical and embodied knowledge that is deemed to be relevant to specific phenomenon, is therefore dependent on their social context and on their participation in it. I shall illustrate precisely how this works in practice in chapter 7, where I discuss how the organisation of Arabidopsis research as centralised big science affects the ways in which Arabidopsis biologists think and talk about their results, as well as their experimental practices and their ability to access and manipulate models and instruments useful to further their understanding of plant biology.

2.3.3 An Internalist Position?

I would like to spend some final words on an objection that is likely to arise in relation to my – for now – minimalist sketch of what it is that enables scientific understanding. This is the idea that my approach is overly demanding: it sets high standards for the set of tools and conditions necessary for an individual to understand phenomena 'scientifically'. As I shall clarify in the rest of this dissertation, there are different ways in which understanding can be achieved: biologists use several strategies to coordinate their theoretical and embodied knowledge, depending on the type and quality of the knowledge that they possess, the phenomena with which they are dealing and the social, institutional and material setting of their research. Still, my definition of scientific understanding might be criticised as strongly internalistic[51] - in other words, as restricting the ability to understand scientifically to active participants in research communities, thus denying that people without scientific training and professional experience might be able to understand in such a manner.

There is a sense in which this criticism is well taken. Indeed, I believe scientific understanding to be strongly tied to participation in scientific communities and related exposure to their epistemic culture(s). I have already emphasised how the ability to understand scientifically, as I describe it, is nurtured and regulated by participation in a research community, whose epistemic culture determines the adequacy of the theoretical and embodied knowledge used by the individual in order to understand the phenomenon in question. When saying this, I do not wish to argue that other, non-scientific kinds of

[51] This notion of internalism (which has no relation to the one employed as the opposite of externalism within general epistemology) has been forcefully used by Fuller (2000) to criticise the Kuhnian approach to the philosophy and history of science. Kuhn himself presented his work as an internal approach to history (1970, 209). Fuller and others have long questioned the links between such a view and the social authority granted to science and technology in the public sphere (for instances of such debates, see the 1979 issue of 'Kennis & Methode' [in Dutch] edited by Nauta and de Vries and the 2003 issue of 'Social Epistemology' edited by Gattei).

expertise should be seen as irrelevant to scientific understanding. However, their usefulness towards acquiring understanding dubbed as ‘scientific’ depends on the willingness by scientific peers and circles, to accept this knowledge as relevant to their practices and interpretations of natural phenomena. The idea that understanding can only be dubbed as scientific when recognised as such by scientific experts does not mean to reduce scientific understanding to a matter of social consensus, nor does it imply that understanding that is not acknowledged as scientific has no epistemological value. As I discuss above, the definition of understanding as scientific depends *both* on the manner by which it is achieved and on the consensus of the community in which it is expressed. Thus, an individual understanding a phenomenon through coordination of relevant theoretical and embodied knowledge is producing valuable insight: its relevance to science is, however, dependent on whether and when a scientific community adopts such understanding as its own.

This account does not imply by any means that non-scientists cannot or should not intervene in scientific controversies: this involvement is both possible and necessary within societies that are increasingly dependent on scientific and technological developments. What my account of scientific understanding does signal is a sizable gap in resources, perspectives and expertises between scientists and non-scientists. This gap is a core issue underlying the democratic use of science and technology (also referred to as ‘politics of displacement’; Beck 1991) and it is my opinion that it should be recognised as such. Let me underscore this view by appealing, again, to the case of GMOs. That the results of research on GMOs are difficult to understand in the absence of detailed knowledge about transgenic technology and genetic mutations, is evident in the civil unrest surrounding the introduction of GMOs in the European Union. Several representatives of European civil society have been expressing deep concern about their inability to assess the risks eventually associated with GMO consumption. The voicing of these concerns effectively hampered ongoing research on GMOs (especially on those destined to human consumption, such as maize, tomatoes and apples). Arguably, a prominent cause for consumers’ worries consists precisely in the knowledge gap separating them from biologists, which makes consumers unable to understand the biological processes involved in biotechnological interventions. As a consequence, it becomes more difficult for non-scientists to evaluate health risks that might be associated to the consumption of such organisms. Several other controversies surrounding the use of technologies (such as the environmental hazards caused by chlorofluorocarbons [CFCs], long-term effects of radiations emitted by mobile phones and other house appliances, and so on) signal the same type of unease: debates among scientists on the safety of these technologies are just as heated as debates on the same topic by non-scientists (and the divergences on what is considered to be ‘scientific’ might be large in both camps), but the topics discussed and especially the means by which discussions are carried out tend to differ considerably between the two spheres. While using a technology does not usually require theoretical knowledge of the processes underlying its functioning, such theoretical knowledge, together with embodied knowledge of how scientists apply theoretical knowledge in producing an artefact, is required in order to understand how that artefact works and with which long-term implications. This latter type of scientific understanding is often crucial to assessing the impact of that artefact on human health and the environment.

In this thesis, I do not intend to propose a solution to the complex and much-discussed problem of the role and justification for so-called ‘scientific expertise’. However, I find it instructive to signal the connection between my present discussion of the epistemology of scientific understanding and the more political discussion over internalism in science, in the hope that looking at ‘expertise’ from my perspective might elicit further, innovative research on this important set of issues. I shall return to this in my conclusions (section 8.1.1).

Chapter 3. History of Arabidopsis Research: Towards Integrative Plant Biology

3.1 Model Organisms in Biology

> *Reasoning via model organisms, in a sense, has become the lingua franca of biologists entering twentieth-first century*
>
> *Rachel Ankeny 2001, S251*

The selection of appropriate organisms on which to make experiments and from which to draw data has been a pressing problem throughout the history of biology. Biological research typically aims at extrapolating knowledge whose applicability extends beyond the organisms that are actually being studied. The study of an individual organism is taken to provide understanding about all other members of the same species; further, it is often expected that the study of a single species should provide biological insight into many other species. This idea is grounded in evolutionary theory, according to which all life forms are related through a common evolutionary history and thus share a smaller or greater amount of genetic make-up and developmental features. This assumption is used to justify the possibility that an organism acts as a 'sample' of a much larger class of animals and/or plants (those, that is, that are phylogenetically most closely related to that species and hence display significant morphological, structural or ecological similarities with its members).

This possibility turns into a necessity when, for ethical and/or pragmatic reasons, research cannot be carried out directly on the organism of interest – blatant examples being many human diseases and relevant treatments, which cannot be investigated without serious hazards to the health of the organism that is employed. An important cluster of issues surrounding the use of model organisms in biological research concerns precisely its ethical and scientific soundness, especially with regard to pharmaceutical testing. Is it at all viable to propose that knowledge acquired on rats is safely transferable to humans, and if so, under which conditions? Is it possible to formulate criteria through which the representativeness of a model organism can be evaluated? There is no easy answer to these questions and, while they are not entirely irrelevant to my discussion, it is not my purpose to confront them in this dissertation.[52] What interests me here is the *ubiquity* and *epistemological significance* of model organism research in biology, as there is simply no way, in practice, to avoid (1) basing research on the study of actual organisms and (2) treating them as representatives for other organisms. As Krogh concluded already in 1929, 'for a large number of problems there will be some animal of choice or a few such animals on which they can be most conveniently studied' (1929, 244). This is true even

[52] For instructive discussions of the ethical dimensions of model organism research, see Lynch (1988), Logan (2001, 2002) and Malezka et al. (1988). Some relevant literature is restricted to animal experimentation, e.g. Rowan (1984) and Lafollette and Shanks (1997).

despite the popularity recently acquired by simulations and digital models in experimental biology (in fact, as I shall discuss in chapter 6, research via these types of models yields useful insights especially when complemented by research on actual organisms).

The central role played by model organisms in biology is due to their crucial contributions towards scientists' understanding of biological phenomena. I shall clarify precisely how this is the case in Chapters 6 and 8 of this dissertation. For now, I take this claim to imply that the use of model organisms has several *epistemological advantages*, some of which I shall introduce and discuss in this chapter. Model organisms are often chosen on the basis of their tractability[53] under laboratory conditions, including ease of maintenance and transport, length of life cycles and fertility rates. They provide a stable environment in which a biologist can investigate specific issues (such as the animal's metabolic system), while leaving worries about environmental variability across individuals or about variation among species aside for later consideration. They thus constitute low-cost, low-maintenance research materials that are easy to control and of which a substantial body of knowledge can rapidly be accumulated, since repeated use of and reference to the same model organism provides a great opportunity for sharing knowledge across the vast constellation of biological disciplines, groups and research schools. As I discuss in the chapters that follow, the choice of a model organism greatly influences the way in which biologists reason about the issues that they investigate, the way in which they act in the laboratory (for instance, by imposing protocols and skills favouring the survival and maintenance of the organism at hand) and the way in which they interpret such actions.

The profound impact of model organism research has been amply recognised by biologists themselves, who have exploited model organisms in increasingly sophisticated manners and have recently magnified their usefulness as research tools by giving them central stage in the genomic revolution. By the end of the last century, the transparent nematode *Caenorhabditis elegans*, the bacterium *E. coli* and the plant Arabidopsis had become the first organisms to have their DNA fully sequenced, thus opening a whole new space for genomic research aimed at deciphering the now cracked, and yet largely uninterpreted, code. As a consequence of their new status in an ever-expanding microbiological research, the *use* of these model organisms has been made more efficient - for instance by creating and standardising protocols and codes of conduct in order to keep them alive and handle them adequately, as well as procedures and techniques to make them 'cooperate' with the other human and non-human actors in an experimental setting (Latour, 1987). Biologists have also found ways to enhance their *production*, so as to obtain and reproduce organisms in the shapes and/or genetic make-up that are most desirable for specific research programmes. Last but not least, the *distribution* of model organism has been standardised: various model organisms communities, including the

[53] Note that the term tractability, as I use it in this dissertation, indicates the ease with which specimens can survive and reproduce in a laboratory. It does not indicate the ease with which specimens can be modified to fit the techniques and procedures of specific types of research (as for instance in the case of mutations).

Arabidopsis community, have instituted stock centres and private inventories that make them available to laboratories that require them.

An important condition for, as well as implication of, these processes is that *large research communities* have been forming around the use of specific model organisms, so as to profit from the know-how, expertise, instrumentation and data accumulated by participating scientists. Funding programmes have increasingly supported work on the most popular model organisms, thus creating the so-called 'founders effect' in biological research (that is, the narrowing of experimentation to a few well-studied organisms, rather than bestowing resources on comparative research among organisms) as well as a complex network of alliances and interests gravitating around and within the communities devoted to one or the other organism. Glaring examples of this tendency are the fruit fly *Drosophila melanogaster*, whose claims to fame stem from its adoption as *the* model organism for genetics – a choice so successful as to promote the humble insect to 'lord of the flies' and centre of attraction for virtually all researchers in the field (Kohler 1994); the above-mentioned worm *C. elegans*, 'conqueror' of neurology (Ankeny 1997, 2000, 2006); the crucial contribution of the tobacco mosaic virus to the development of molecular biology and pathogeny (Creager 2002); and the house mouse, whose intricate adventures as the object of study of several biomedical branches have been brought together in a recent volume by Karen Rader (2004).[54]

I shall not attempt here to summarise the various ways in which reference to a specific model organism impacts scientific reasoning: it is my intention to explore them in detail throughout this dissertation (particularly in chapter 6), with reference to the case of the model plant *Arabidopsis thaliana*. Arabidopsis research is propelling significant advances in several branches of the life sciences. The success story of this plant sets the stage for the epistemological analysis of the main features of model organism research that I will propose in Chapters 5, 6 and 7. My aim in this chapter is to unravel at least some of the factors that contribute to the success of a specific life form as a model organism. To this aim, I start from an historical reconstruction of the events that transformed Arabidopsis into the leading model organism in plant biology. With reference to this history, I shall single out and discuss several characteristics of Arabidopsis research that account for its scientific fruitfulness and institutional fortunes. This analysis constitutes an appropriate starting point for my philosophical study of how practices in model organism research enable a biological understanding of the phenomena under study.

At the same time, the very nature of these characteristics, which include both social and scientific factors, colourfully illustrates the importance of collective work in biological research. In line with research modes characterising other large model organism communities, the Arabidopsis community is engaged in big science, in which sponsors,

[54] This list is by no means exhaustive. For other examples, consider the literature on rats (Logan 2001), yeast (Botstein and Fink, 1988), further studies of C. Elegans (Schaffner 2000, de Chardaverian 1998) and all examples cited by Geison and Creager (1999).

participants, projects, investments and (it is expected) results all come in high numbers. Its scientific agenda, activities and results are therefore shaped by a variety of factors, including power dynamics, social tensions within and without the community, political commitments, economic constraints and scientific alliances. As claimed by historian Angela Creager, 'the uses of model objects in research reveal the otherwise inconspicuous connections between biological experimentation and activities usually relegated to the domains of technology, politics, medicine and agriculture' (2002, 3). I hope to illustrate how such connections inform and maintain the epistemological status of Arabidopsis research. The collective nature of this enterprise, as well as the many non-scientific factors that make it possible, greatly influence the way in which scientists conceive of a model organism and manipulate it in order to gain knowledge about biological phenomena. As a prelude to further discussion of this point, I shall conclude this chapter by introducing two subgroups in the Arabidopsis community, the details of whose work shall inform much of my analysis: The Arabidopsis Information Resource and the Nottingham Arabidopsis Stock Centre.

3.2 Enter *Arabidopsis thaliana*

> *One of the most powerful forces in plant biology during the past 15 years was the accumulation of a critical mass of scientists around the use of the small plant Arabidopsis thaliana. Much of the recent progress in plant biology has been made possible by the technical advantages of having a large group of biologists working with a well-chosen model organism combined with the inherent power of molecular genetics*
>
> *Chris Somerville 2000, 21*

Arabidopsis thaliana, commonly referred to as mouse cress, is a plant of the mustard family that grows almost everywhere across the Northern hemisphere. An occasional visitor wandering through the European countryside is, however, highly unlikely to pay any attention to its characteristic white flowers – and with good reason. Arabidopsis is tiny, with an average length of about 10 to 15 cm excluding the roots. It has a short life-span (about six to eight weeks), bears no fruits and its flowers come in no spectacular colours, size or variety. As most other types of weed, it is regarded as a parasite, yet its impact on the ecosystem is so limited as to be generally disregarded. It obviously has no agronomic significance. In short, Arabidopsis has long been regarded as an insignificant organism from the commercial, aesthetic as well as scientific points of view.

Figure 3.1 - Arabidopsis thaliana.

This connotation accounts for the very scarce attention that experimentalists and naturalists alike, with a few exceptions in the Netherlands, Germany and the States, paid to Arabidopsis until the 1980s. Yet, over the last two decades, Arabidopsis came to play a prominent role in experimental plant biology: it is currently its most important and well-researched model organism. Since the late 1970s, the community of biologists working with this plant has grown from around 25 researchers to 6000 laboratories distributed in five continents, for a total of more than 16000 researchers.[55] Its genome has been the first plant genome to be sequenced and the facilities and tools assembled to gather data about it are arguably the best available for any organism, with the exclusion of Homo sapiens. Further, this extensive gathering of knowledge has brought about several important theoretical discoveries in a variety of fields in plant biology (table 3.1): among them, an improved understanding of molecular mechanisms for flowering and root development, light reception, metabolism and disease resistance (plant pathogen interactions).

Table 3.1. Most cited subject areas, illustrating the scope of Arabidopsis research. Courtesy of James Evans 2006.

Subject Area	*N* Cites [a]	*N* Papers [b]	Papers / Cites	First Paper
Plant Sciences	115,066	6,512	17.67	1974
Biochemistry & Molecular Biology	92,270	4,319	21.36	1974
Multidisciplinary Sciences	45,257	823	54.99	1974
Cell Biology	25,503	1,577	16.17	1981
Genetics & Heredity	23,669	1,218	19.43	1974
Biology	18,518	684	27.07	1979
Developmental Biology	12,419	409	30.36	1979
Biophysics	5,651	494	11.44	1980
Cytology & Histology	4,166	33	126.24	1974
Biotechnology & Applied Microbiology	4,121	416	9.91	1987
Physiology	713	53	13.45	1977
Biology Miscellaneous	710	85	8.35	1990
Ecology	620	57	10.88	1983
Microbiology	567	61	9.30	1992
Biochemical Research Methods	408	89	4.58	1996
Virology	369	35	10.54	1993
Agriculture	345	78	4.42	1985
Agriculture, Dairy & Animal Science	278	30	9.27	1990
Chemistry, Analytical	269	28	9.61	1988
Toxicology	253	28	9.04	1979
Reproductive Biology	249	30	8.30	1996
Reproductive Systems	218	8	27.25	1990
Biomethods	186	8	23.25	1990
Agriculture, Soil Science	155	34	4.56	1991
Botany	148	9	16.44	1976

[1] Subject areas were developed by the Institute for Scientific Information (ISI), and are represented only for the 13,000 *Arabidopsis* articles found in ISI's. database.
[a] *N* Cites refers to the total number of scientific citations to *Arabidopsis* articles published within these subject areas.
[b] *N* Papers refers to the total number of *Arabidopsis* papers published within this subject areas.

[55] This estimate is based on the number of contributing members registered by the The Arabidopsis Information Resource (TAIR website, accessed 15 November 2006), as well as from statistics gathered from published Arabidopsis research by the sociologist James Evans (2004, 2006).

This scenario prompts several questions, the most prominent of which for my current purposes is an historical one: how to account for this reawakening of interest in Arabidopsis and, most importantly, for the ensuing success story of the plant as a model organism? This is the question that I shall address in this section of the chapter. To this aim, I shall first review the history of Arabidopsis as a model organism, then focus more closely on the features of the plant that, in the context of a specific community and scientific setting, have determined its unlikely success.

3.2.1 Early History of Arabidopsis Research

The first significant bout of interest in Arabidopsis as a model organism arose in Germany and the United States in the 1940s, mainly through the work of Friedrich Laibach (based in Frankfurt) and George Redei (Columbia). Laibach was initially attracted to Arabidopsis because of the great phenotypic variation to be found in nature among its ecotypes: he was in fact the first scientist, in 1937, to start a systematic collection and classification of Arabidopsis wildtype mutants. This interest was fuelled by his belief that Arabidopsis could become a suitable model organism for (classical) plant genetics and developmental research, i.e. for the mechanisms responsible for this diversity of ecotypes (Laibach, 1943). Several material features of Arabidopsis made it a promising candidate for the experimental study of natural variation at both the ecological and the micro-structural levels. First, there is its ease of maintenance and small space requirements, allowing the cultivation of large populations of plants within relatively primitive facilities. Second, its low chromosome number – only five, which is little relative to other flowering plants. The last and most important element is its short generation time and large number of progeny: a self-fertilising diploid, each Arabidopsis specimen can produce thousands of seeds in its short life cycle (between six and eight weeks).

Laibach's collection of mutants reflected his interest in both the ecological and micro-structural properties of Arabidopsis. Working in the Mendelian framework, he was interested in Arabidopsis ecotypes that presented variations at the phenotypic level, in order to study and compare their differences at the chromosomal level and thus establish correspondences between chromosomal and morphological traits. Laibach's intuitions were developed in the 1950s and 1960s by research groups led by Michel Jacobs in Belgium, Wil Feenstra in the Netherlands and Gerhard Röbbelen in Germany. Upon Laibach's retirement in 1965, Röbbelen assumed the role of curator of the ecotype collection (which he shared with Redei in the United States, so as to have a set of all lines per continent). At around the same time (1964), Röbbelen also started the publication of the Arabidopsis Information Service [AIS], a yearly newsletter bringing together research updates on experimental work on the plant and helping communication and exchanges among interested biologists.

The small group of readers and contributors gathered by AIS was not yet, however, destined to grow. In fact, despite the convincing evidence available about the utility of Arabidopsis to the study of genetics, these early attempts to bring Arabidopsis to the laboratory failed to create wider consensus. This was mainly due to the values and interests upheld in plant biology as a whole in the post-war period. While huge amounts of money were destined to basic genetic research on animal models, both in Europe and in the United States most governmental (not to mention corporate) support for plant biology was directed towards applied research. Projects concerning breeding techniques, particularly on agriculturally significant organisms such as tobacco, received almost all of the anyway limited funding allotted to this branch of biology. In the absence of institutional support, basic research on commercially uninteresting plants, such as 'insignificant' Arabidopsis, remained highly unfashionable among plant biologists. Moreover, a second major drawback emerged from the study of the material properties of the plant. While its natural variability attracted biologists working in the framework of classical genetics, the new emphasis on biochemically induced mutations proved to be inapplicable to Arabidopsis specimens: they simply refused to be transformed. The few mutations obtained between 1950 and the early 1980s were produced at the high cost of months of experimental labour and with no foreseeable hope of speeding up the process. Arabidopsis' resilience to genetic modifications was an extremely unhelpful feature at such a high point of the molecular bandwagon: upon realising this, the few European and American biologists hitherto interested in working on Arabidopsis had to turn the bulk of their energies to tomato, yeast or maize. Even Maarten Koornneef, a student of Feenstra in Wageningen and one of the early and most fervent champions of Arabidopsis as a model system for genetics, retreated to the study of tobacco and published on Arabidopsis in his (very scarce) spare time.[56]

In the face of this debacle, how can we explain the revival of Arabidopsis as a model organism that blossomed already at the beginning of the 1980s? A first step towards understanding requires enlarging our historiographic gaze to include the contemporary situation in animal biology, and particularly the burgeoning field of animal genetics. Many researchers working on Drosophila felt that the time was ripe for moving genetic analysis decidedly in the direction of molecular biology. However, the large resources, long history and high reputation of research on fruit flies constituted obstacles, rather than advantages, for this purpose. Competition among the several researchers enrolling in the Drosophila community was stiff and unforgiving: as a result, newcomers to the field were encouraged to specialise on already established – and thus often theoretically uninteresting – issues. Exploring new directions of research in such a climate was almost impossible, given both time and peer pressures to submit enough publications as required to climbing the steep academic ladder. Plant biology, by comparison, was regarded by

[56] Other factors also contributed to the increasing disinterest in Arabidopsis as a model organism. For instance, there was the interest initially awakened by the publication of a paper trying to force Arabidopsis to assimilate thiamine by injecting E. coli DNA into the plant (Ledoux, L., Huart, R. and Jacobs, M. 1974). The experiment proved to be entirely spurious, thus further endangering Arabidopsis' shaky reputation as a model organism. The rise of scepticism following this publication is documented in Somerville and Koornneef (2002) and has been further emphasised to me by Koornneef himself (pers. com.). Further, personal misfortunes also contributed to Arabidopsis decline. To mention one: a PhD student hired by Feenstra in the mid-1970s to pursue his interests in biochemical genetics for nitrites metabolism in Arabidopsis, was diagnosed with multiple sclerosis shortly after her appointment. Human tragedy thus also triggered a significant delay in (in this case, Dutch) genomic research on the plant (since the student, who indeed managed to conclude her thesis, submitted her results only well into the 1980s).

many geneticists as a relatively 'empty field' (Koornneef, pers. com.). Focused as it was on producing applied knowledge (such as breeding studies), the discipline did not have a history of tackling basic developmental mechanisms in plants. Here, therefore, was a possibility to start this line of research anew in a much less competitive environment – which meant, on one hand, freedom from the cumbersome legacy of the theoretical commitments favoured by older biologists; on the other hand, the possibility to organise research effort in a highly collaborative fashion, so that participant scientists would investigate different and potentially complementary topics on a common basis of insights and tools coming from molecular biology.

Given the situation in animal molecular biology, the dormant state of Arabidopsis research characterising the end of the 1970s turned into a contributing factor for its revival. Precisely when animal researchers started to feel that 'molecular biology was in such a state that it needed a model' (Koornneef, pers. com.), they discovered a plant of excellent tractability and chromosomal structure, on which a host of genetic data was already available together with a rich and well-systematised collection of ecotypes, yet which was not already adopted by any large research project. Some prominent experts in animal biology, such as Elliott Meyerowitz (Caltech) and Gerd Jürgends (Tübingen), thus started to consider channelling their interest in organismic development towards plants. Even more enthusiastically, younger researchers of great promise understood the situation and decided to gamble their future careers on it: prominent among those figured Chris Somerville and Shauna Somerville, who teamed up (in research as well as in private life) to start one of the first and most successful programmes in Arabidopsis research first at Michigan State University and then in Stanford's Carnegie Institution for Plant Biology.[57] In their own words, the relevant question for molecular biologists in the early 1980s became: 'why *not* a plant?' (pers. com.).

Figure 3.2 – First meeting of (from left to right) Shauna Somerville, Chris Somerville, Elliot Meyerowitz, David Meinke, an unidentified researcher and Maarten Koornneef. Photograph: courtesy of Chris Somerville and Elliot Meyerowitz.

[57] The vision of the Somervilles played indeed a crucial role towards the establishment of the Arabidopsis community. During an interview conducted in August 2004, they recounted to me what they felt to be the circumstances in which they conceived their ideas about the future of Arabidopsis. Right after Shauna's graduation from her PhD on Arabidopsis pathology, she and Chris (then working on E coli after graduating in Maths and shifting to experimental biology) left for a summer vacation in Paris. There, they apparently spent their time elaborating a life-career plan for themselves based on their specific vision about what could be achieved via Arabidopsis work. Shauna convinced Chris that work on that plant held great promise, both for its characteristics as a model organism and because of the humanitarian promise held by biotechnology. Notably, the Somervilles believed that work on plant biology could have positive, long-term impacts on people's lives (via the improvement of control over plant development which, when applied, would lead to enhancements in agricultural productivity).

It was to discuss this question and its possible implications that the Somervilles, Koornneef, Meyerowitz and David Meinke (now at Oklahoma State) independently decided to attend a 1985 meeting on plant molecular biology in Colorado. That otherwise unremarkable conference had the important function of bringing together for the first time the founding fathers of the 'revived' Arabidopsis community. The group hit off immediately, as their communality of interests, values and vision quickly transformed their affiliation into a friendship (figure 3.2). This is an important element in the history of Arabidopsis as a model organism, since the strong personal and professional bond established among its advocates greatly helped to enhance the popularity of the plant among other researchers. This was in part due to the group's complicity in elaborating common strategies towards attracting funding (and thus recruits and facilities) at both the national and the international level. Even more importantly, the group expressed a strong commitment to creating a common ethos for the prospective community of Arabidopsis researchers: it was an ethos of collaboration and coordination of research efforts, aimed at heightening the overall output of the community as well as countering the 'publish or perish' mentality increasingly pervading the scientific world.[58] Sue Rhee, a former

[58] Remarkably, both Chris Somerville and Koornneef have emphasised to me how struck they were, since their youth, by the epistemic advantages of a community sharing data on the basis of need and interest, rather than as a career token. Koornneef, for instance, had an early experience of collaborative research when visiting Zürich as a postdoctoral fellow (then a recognised centre of excellence on plant biology research) – an experience that had a dramatic impact on his perspective on research methods and values. Chris and Shauna Somerville, as a married couple working on different aspects of the plant, also have extensive personal experience in collaboration. Their current graduate students at Carnegie do not even understand the competitive attitudes that they witness outside Arabidopsis research (indeed, even now the

student of Chris Somerville and current director of The Arabidopsis Information Resource, aptly characterised this attitude as the 'share and survive model' of research (Rhee 2004). In 1987, an international steering committee called Multinational Arabidopsis Steering Committee [MASC] was created in order to coordinate Arabidopsis research projects around the world.[59] To this day, MASC functions both as a powerful attractor for governmental funding and as a centralising and rationalising force in the community, subtly but strongly enforcing the ethos of collaboration and pre-publication sharing of results that would have otherwise been lost to the competitive dynamics of scientific hierarchies.

In plant biology, the growing attention to molecular approaches and the increasingly successful campaign in favour of Arabidopsis research brought about a shift in the features to be valued in a model organism. It became acceptable to study a plant that was highly suited to laboratory life, but of no immediate agronomic interest, since the basic mechanisms of heredity were expected to be applicable to all other plants, including more useful crops. This was particularly true in the case of Arabidopsis, which is genetically closely related to other flowering plants and is thus characterised by a high (genetic) typicality with respect to them. These characteristics partially account for the choice of Arabidopsis as *the* plant on which to focus research effort. A major remaining obstacle to its adoption was, however its above-mentioned low transformative ability. This feature nurtured much of the lingering scepticism about focusing on plants, in general, and Arabidopsis, in particular, as models for molecular genetics. No wonder, then, that the unexpected announcement of a spectacular solution to this problem, published by Feldmann in 1986, felt like 'hocus pocus' (Chris Somerville, pers. com.), a little 'miracle' (Koornneef, pers. com), to early Arabidopsis proselytes. What Feldmann discovered was a simple and extremely effective technique for the production of Arabidopsis mutants: it was sufficient to spray the wildtype with a bacterium (*Agrobacterium tumefaciens*) containing a gene of interest in plasmid, then selecting for the transgene in the second generation (Feldmann and Marks, 1987). Thanks to this technique, it became suddenly possible to obtain an astonishing variety of Arabidopsis mutants in which phenotypic growth had been disrupted in some way or other. This development won a great number of biologists over to the study of Arabidopsis.

Last but not least, there was the sudden decision by the National Science Foundation to appropriate the study of the plant and provide abundant funding to its American proponents. The interest and lobbying of James Watson, by then a celebrated figure with powerful influence on governmental support for scientific projects, certainly contributed to the benevolence of NSF.[60] Another strong motivation for NSF involvement had again to do with the state of research on animal models, which was then entirely funded by the National Institute of Health [NIH] – a governmental agency traditionally competing with

---

collaborative ethos characterising this community represents strong motivation for young researchers not to move to other organisms). More on this in Chapter 7, section 7.1.1.

[59] While its members are elected by open nomination and voting, and their mandate lasts no more than 3 years, the MASC committee tends to include researchers who have studied with one of the 'founding fathers' or who anyway subscribe to their vision and long-term goals.

[60] Watson's involvement was indicated to me by Chris Somerville (pers. com.).

NSF for prestige and power. NSF hoped – and indeed, successfully managed – to exploit the empty niche of plant biology in order to enhance its own profile among American funding bodies. Its institutional support was crucial not only to the rapidity with which Arabidopsis research developed in the States, but also to the manner of that development, since dependence from a unique funding source enormously enhanced the centralising and collaborative nature of the Arabidopsis community. Also, NSF patronage greatly spurred governmental support at the international level, since none of the other big players in scientific research could afford being left behind. In five years from the start of NSF funding, therefore, Britain, Germany, the European Union and Japan had all agreed to invest considerable amounts of money into Arabidopsis research. Further, they had to yield to the ethos of collaboration and sharing of resources imposed by their American partners: it quickly proved unthinkable to keep secrecy over results at the pre-publication stage within an American-dominated context supporting that ethos.[61]

3.2.2 Creating a 'Botanical Drosophila'

> *Standard flies were not just the means of experimental production but also the bearers of a distinctive moral economy and a distinctive way of experimental life*
>
> *Robert E. Kohler 1994, 168*

All the above-mentioned factors, ranging from the useful features of Arabidopsis plants to the institutional setting and resources devoted to their study, resulted in the birth, in 1990, of the Arabidopsis Genome Initiative [AGI], a multinational research effort (involving six research groups based in the States, Japan and Europe) that yielded a complete mapping of the Arabidopsis genome by the year 2000. As the first plant undergoing genomic sequencing, the status of Arabidopsis in plant biology was finally sanctioned: in little more than a decade, it had become its best-known and most popular model organism. This latter statement refers to the institutional as much as the epistemological status of Arabidopsis in plant biology. The recognition of a model organism is taken here to involve both the frequency of its usage in the laboratory and the type of knowledge yielded through that usage. In other words, the dramatic growth in the number of researchers working on Arabidopsis contributed to transforming the plant into the most popular model system in plant biology and one of the most important organisms in biology as a whole.

This argument implies that the extent to which a model organism is used does tell us a lot about how it is used and to which results. That is, the study of how a research tool is used by scientists is relevant towards uncovering its epistemic function and usefulness. The history of Arabidopsis as a research tool and (as I will characterise it in Chapter 6) a materially abstracted model constitutes a clear illustration of how closely social,

[61] Sean May, director of NASC, stressed the initial British resistance to this ethos in his interviews with me. The US 'push' is also visible in episodes such as the Crete meeting (early 1990s) organized by the EU to select which model organisms to privilege within plant biology. On that occasion, both the EU and Japan stressed how they had to follow the US trend and fund Arabidopsis research.

institutional and scientific factors intertwine in order to enhance and shape the epistemological contribution of Arabidopsis to plant biology. As reported above, many elements of the early Arabidopsis history signal the importance of institutional support, community ethos and strategies to obtain funding in shaping the goals of research on Arabidopsis. In fact, the centralised coordination of Arabidopsis research was guided by the enthusiastic acceptance, by participant researchers and funding agencies alike, of a common scientific goal concerning precisely the status of Arabidopsis plants as model organisms. This goal, as expressed by the first MASC programmatic statement in 1990, was 'to understand the physiology, biochemistry, growth and development of a flowering plant at the molecular level, using Arabidopsis as an experimental model system' (MASC 1990, 12). Remarkably, this formulation implies that the Arabidopsis community as a whole takes up a commitment for which it holds itself accountable to its sponsors and peers: this is to treat Arabidopsis as *a model of the molecular processes* characterising growth and development in flowering plants.

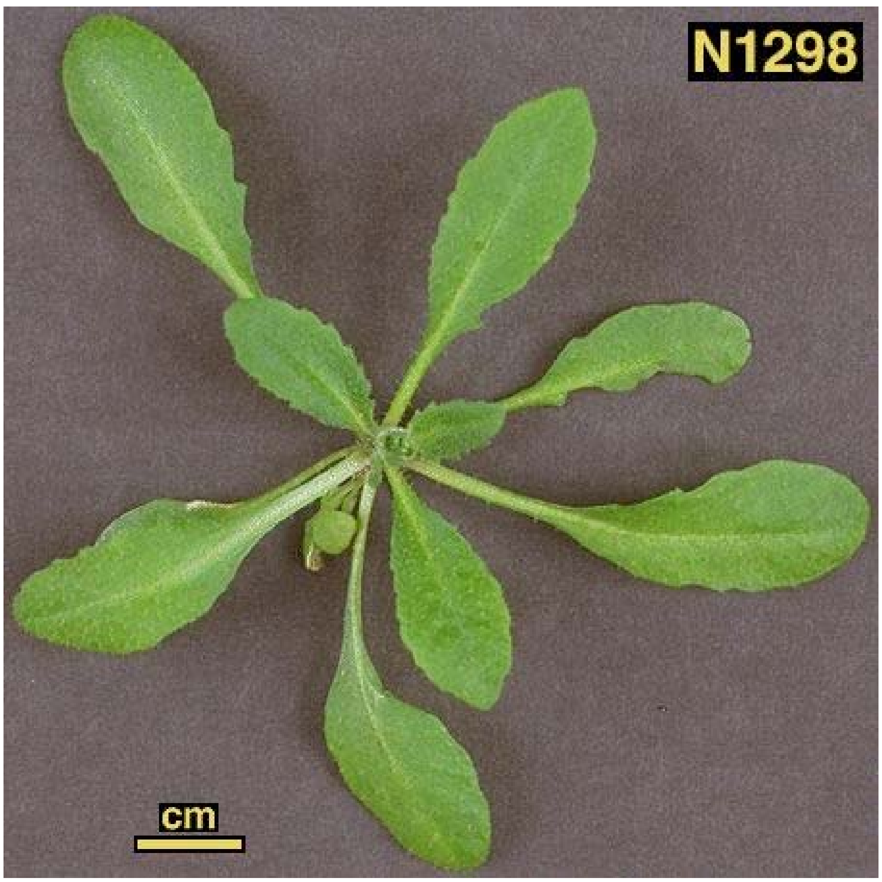
As indicated by the disciplinary background and research interests of most scientists involved in the community up to the early 1990s, what confirmed the importance of Arabidopsis as a model system was indeed its relevance to the study of plant genetics and molecular biology. Already in 1964, the first AIS editorial exemplifies the extent to which this goal set an agenda for Arabidopsis research. In it, Röbbelen characterised the plant as the 'botanical Drosophila', that is to say as an organism useful to the study of the biochemical basis for genetic mutation, in the same way as the celebrated fruit fly. This was not surprising in light of the paradigm shift to molecular genetics, or 'molecular bandwagon', characterising a large part of biology at the time.[62] Yet, it also indicated the gradual dismissal of Laibach's original interest in the morphology, evolution and ecology of Arabidopsis as a necessary complement to its genetics. Despite Röbbelen's feeble reminders to pursue and publish research concerning other aspects of plant biology[63], the AIS became a promotional vehicle for shifting research focus from the phenotypic to the genotypic features of Arabidopsis. This communication tool exemplifies the extent to which the adherence of institutions and researchers alike to specific epistemic commitments facilitated the rise of Arabidopsis as a credible system for the study of mutagenesis. By the mid-1990s, most scientific reviews hailed the usefulness of Arabidopsis as 'a model plant for genome analysis' (Meinke et al, 1998): in other words, it was not the plant as a whole organism that counted as experimental tool, but rather its chromosomes.

It seems evident that the institutional set-up and centralised ethos of the early Arabidopsis community were indispensable conditions for the establishment of a *common* scientific goal for the use of the plants. However, this is not the only way in which the research community made it possible to treat Arabidopsis as the 'botanical Drosophila'. A key contribution to the achievement of this goal was the shift in the administration of the stock of Arabidopsis specimens originally collected and systematised by Laibach. Between 1950 and the 1970s, Laibach's collection was maintained and moderately

[62] For historical accounts of this shift, see Kay (2000) and Keller (2000).
[63] 'Genetics will indeed take the main part, but notes on ecology, morphology, development, physiology or biochemistry should as well be appreciated' (Röbbelen 1964, 1).

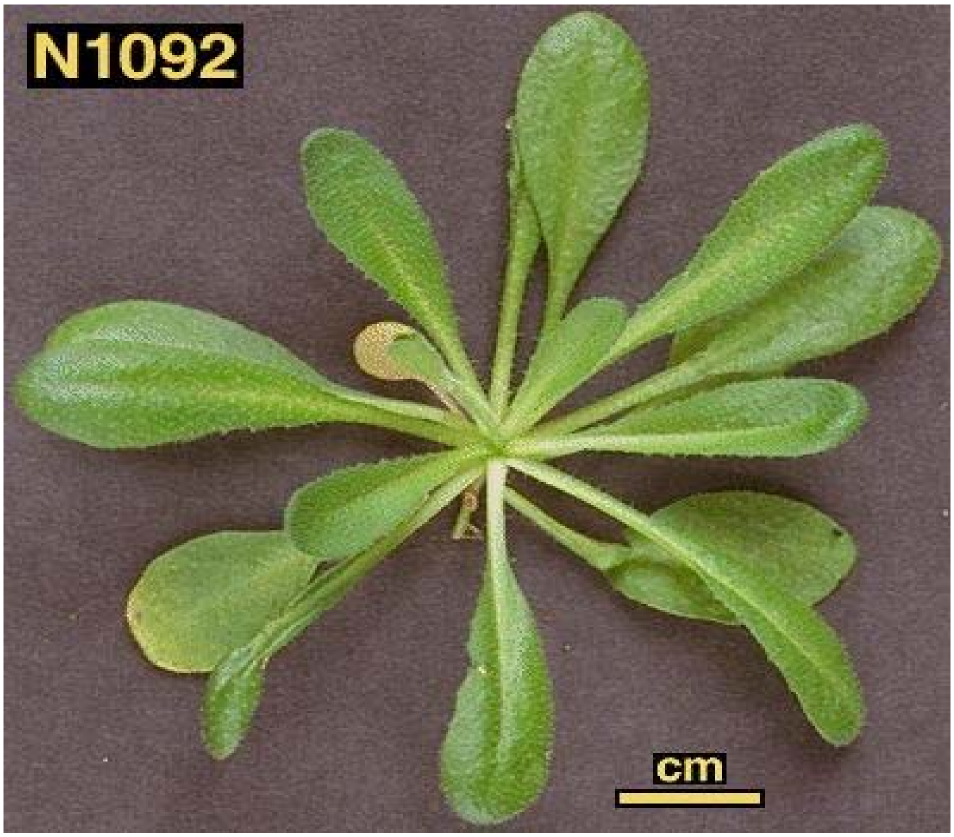
enlarged by Redei in the States and by Röbbelen and Albert Kranz in Germany. The German collection, institutionalised in 1965 into the first official Arabidopsis stock centre, underwent moderate extensions by incorporating Kranz's own samples, but otherwise remained the main point of reference for Arabidopsis researchers up until the late 1980s – that is, until Feldman started to produce mutant lines on a massive scale. The lack of interest in adding ecotypes to the collection is in fact an expression of the representational value attributed to the plants in that period. Researchers were not interested in the extent of morphological variability among Arabidopsis specimens in the same way as Laibach was at the beginning of the century: they wanted to investigate the common mechanisms underlying such diversity. The research spotlight was thus placed largely on only two ecotypes. The first was a wildtype originally classified by Laibach under the name of Landsberg [Lan] (figure 3.3).

Figure 3.3. The Landsberg ecotype.



By 1970, the Lan wildtype had undergone already a considerable amount of modifications for use in the laboratory. Redei had selected it out of Laibach's collection and irradiated it with X-rays, in order to derive genotypically identifiable populations. One of those was adopted by Feenstra under the name of Landsberg *erecta* and became known as the background line for all the mutants produced by his group at Wageningen (including Koornneef). The second chosen ecotype was a particularly fertile and vigorous strand which Redei had selected from the original Landsberg nonirradiated population. It was referred to as Columbia [Col] and acquired enormous popularity especially in the 1990s, when it was chosen as the background line for the AGI sequencing project (figure 3.4).

Figure 3.4. The Col ecotype.



In both cases, the adoption and progressive standardisation of the ecotype in a particularly visible group of research projects, whose results constituted a stepping stone for further experiments, had the effect of encouraging researchers to pursue research on that ecotype rather than others, so as to guarantee continuity and reproducibility of results. Col and Lan are very different morphologically – the former tall and slender, with slim, pointy leaves; the latter short and rich in well-rounded leaves – as well as genotypically (they vary by 50,000 polymorphisms, thus making it relatively easy to isolate mutant genes by positional cloning). In between the two, researchers agreed that there was enough genotypic differentiation as to allow for an in-depth study of genetic mechanisms. In other words, taken together, Col and Lan constituted a satisfactory model system for the study of plant genetics.

Were the epistemic aims motivating the use of Arabidopsis as a ‘botanical Drosophila’ successfully fulfilled? In light of the successful completion of the AGI, as well as several important findings obtained by the mid-1990s about other molecular processes in

Arabidopsis (such as the biochemical mechanisms for light reception and flowering), the answer to this question has to be positive. Col and Lan proved to be an extremely valuable model of plant genomes for molecular biologists. And again, the internal organisation of the community, including its communication tools and the standardisation techniques for Arabidopsis ecotypes, was crucial to enabling researchers to work on shared models yielding knowledge about plants genetics.

3.2.3 How to Choose a Weed

> *For all of its superior properties, Arabidopsis is typical of flowering plants in its morphology, anatomy, growth, development, and environmental responses, a kind of "everyman" of the plant world. In short, Arabidopsis thaliana is a biologist's dream: a model plant*
>
> *National Science Foundation 2000*

The story of the recasting of Arabidopsis as a botanical Drosophila allows me to locate the major characteristics of successful model organism research, which I will now spell out and to whose analysis I will devote the bulk of this dissertation. Let me start by considering a schematic list of factors that have arguably contributed to Arabidopsis' success:

1. *the manipulability of Arabidopsis specimens*: the initial selection of Arabidopsis as a model can be explained pragmatically by the ease with which the plant can be grown and manipulated in laboratory settings, as well as by its conveniently short life cycle;

2. *the small size of Arabidopsis genome*, which makes it amenable to detailed molecular analysis (as demonstrated by Koornneef's genetic map, released in 1983, and Meyerowitz complementary research published in 1985);

3. *the mutation rate of Arabidopsis specimens*, especially following the discovery of Agrobacterium-induced mutations;

4. *the typicality of Arabidopsis*: in its quality of one of the most common and widespread types of plant (a species of weed), Arabidopsis is used as representative of a large group of plants (except, of course, fruit-bearing species);

5. *the usefulness of Arabidopsis research in the fields of physiology, biochemistry and development:* this was an important argument from the scientific viewpoint, since no model organisms had yet been employed for research in such different, and yet highly interdependent, areas (as argued in influential reviews by Somerville 2002, Meinke 1998 and Meyerowitz 2001);

6. *the organisational and scientific skills of the scientists who first adopted Arabidopsis as a laboratory model* (in the 1970s and then most successfully in the late 1980s), including the commitment of researchers such as C. Somerville and Meyerowitz

towards communication, accountability and exchange of available data (entailing the development of bioinformatic tools and protocols to do so);

7. *the ethos of the community* - the initial and sustained agreement on a broadly defined set of shared goals for future research in plant biology;

8. *the long-term dependence upon the same governmental funding agencies*, such as the NSF, which also acts as an incentive to contribute actively to the community (it is often the case that such contributions constitute an important factor in assessing the quality of each individual's research, since the community includes most of one's academic peers).

We can distinguish two broad categories of factors in this list. The first category brings together characteristics of the model organism that facilitate its use towards gathering biological knowledge, while the second category includes characteristics of the community studying the organism.

Prominent in the first category figure the *natural characteristics* of Arabidopsis specimens. These are the features that are found in all wildtypes independently of human manipulation. They make Arabidopsis into an especially tractable organism as well as helping its representativeness with respect to other plants (points 1, 2 and, to an extent, 4). In contrast to these, I call *induced characteristics* those traits of Arabidopsis specimens that have been artificially created (through biotechnology, as in the case of the oncomouse) or induced via human intervention (as in the case of point 3). These characteristics are helpful towards making a model organism especially tuned to the specific goals and methods of the research carried out on it: for example, given current microbiologists' high interest in genetic variability, it is no surprise that induced mutation has become an important factor for preferring Arabidopsis over other model organisms.

A third type of relevant features of the first category includes the *projected characteristics* of Arabidopsis specimens. These are traits that are attributed to the plant by the scientists using it, without however being or becoming material features of the plant. For instance, the advantages of studying Arabidopsis in physiology and development (point 5) do not directly depend on the specific structure of the plants, but rather on the researchers' intention to use Arabidopsis in this context. Arguably, the typicality of Arabidopsis (point 4) is also a factor that is largely not intrinsic to the plants themselves. This claim might appear counter-intuitive. After all, the extent (and plausibility) to which Arabidopsis can be used as a representative for other plant species seem crucial to its use as a model organism, in the sense specified in section 3.1. Does this not imply that the plants themselves should carry features that are especially – intrinsically – representative? There are two reasons to think that, even if this were the case, it constitutes only part of the justification for taking an organism as highly representative of others.[64] One is that tractability, rather than representativeness, has been shown to be the main criterion for selecting model organisms. A second, more important observation is that the representativeness of an organism results just as much from its

[64] These points are discussed, though in a different context, in Ankeny (2006).

adoption by a research community as from properties intrinsic to the organism itself . Biologists' expectations in choosing a model organism fuel the way in which it is described and compared to other organisms, as well as the representational value given to results gained from such research. This point is usefully emphasised also in Creager's history of the tobacco mosaic virus, where she remarks on the use of model organisms as 'exemplars' in the Kuhnian sense: they are instances of specific ways in which organisms can be handled and interpreted, which can be applied to many other instances depending on the interests and motives of the biologists involved (Creager 2002, 7).

My second broad category of factors focuses directly on such interests and motives, by addressing the features of the scientific community that adopts a model organism, rather than the features of the organism itself. In the case of Arabidopsis, these factors range from *social commitments* (point 6) to *institutional structures and organisation* (point 7) and *funding sources* (point 8). I shall expand on each of these points in Chapter 7. Yet, I hope that my reconstruction of the history of Arabidopsis research already shows the relevance of these factors to the experimental value of a model organism. The Arabidopsis community provides a particularly interesting case for this, since, thanks to the above-mentioned circumstances of its birth, it is an unusually centralised and coherent community. While this makes of Arabidopsis an exception rather than the rule for model organism research[65], I believe that studying this specific community allows an even stronger emphasis on the fruitfulness of investing resources towards a centralised organisation for this type of research, while at the same time providing a close-up view of the tensions and problems also associated with this choice (more on this in Chapter 7).

My analysis of factors involved in the choice and successful use of a model organism is summarised in table 3.2 below. Of course, neither the above list nor my categorisation are supposed to include all factors involved in the choice of Arabidopsis or any other model organism. Yet, they provide a good picture of the main reasons behind Arabidopsis' success as a model plant. They also allow me to exclude factors that, despite their potential importance, had no real influence on the huge popularity that the weed was to acquire. One of them is the great natural variability among Arabidopsis ecotypes. This feature played a relevant role in Leibach and Redei's initial adoption of the organism, since their research focused both on the study of environmental adaptations as well as on the study of genotypic variability. However, it became irrelevant in the context of the subsequent revival of Arabidopsis research in the 1980s: Arabidopsis' rise to notoriety is due largely to its role in plant genetics, rather than to its ecological variability. This latter feature has been revalued only recently, when research on Arabidopsis was anyhow an unavoidable passage point for most plant biologists.

[65] The Drosophila community, for instance, is extremely decentralised despite the greater communality characterising the issues that are investigated (Kohler 1994). This is both due to its long history (starting at the turn of the 19^{th} century and involving research carried out independently by groups in the UK, USA and Russia) and to the competitiveness characterising animal genetics (to which I already referred in 3.2.1). The same holds for the research communities gathered around the tobacco mosaic virus and the nematode C. Elegans. For a much expanded discussion of the centralisation characterising the Arabidopsis community, see chapter 7.

Table 3.2 - Factors contributing to the popularity of a model organism among scientists.

Categories	**Type of characteristics**	**Corresponding Factors in Arabidopsis Research**
CHARACTERISTICS OF THE ORGANISM	*Natural*	Manipulability, small sized genome
	Induced	Mutation rate
	Projected	Usefulness in physiology and development, typicality
CHARACTERISTICS OF THE COMMUNITY	*Social Commitment*	Common goals, ethos of open communication
	Institutional Organisation	Unity, centralisation [e.g. MASC]
	Funding Source	Largely governmental funding [e.g. NSF in the USA]

3.3 Arabidopsis Research Today: Modeling to Integrate

3.3.1 Towards Integrative Biology: The Arabidopsis Information Resource [TAIR]

The extent to which the status of Arabidopsis as a model organism depends on the scientific goals upheld by the community working on it becomes even more evident when looking at what happened in the last decade of research. Starting with the mid-1990s, with its genome almost entirely sequenced, a host of open questions concerning its interpretation and a flourishing community ready to tackle those questions, Arabidopsis scientists suddenly found themselves in a strong position from both the institutional and the scientific viewpoint. Arabidopsis had proved its worth as an extremely valuable research tool in plant molecular biology. Further, the amount of data accumulated on the plant encouraged the above-mentioned ‘founder effect’: that is, the tendency by research communities to keep accumulating knowledge on the same organism in the hope to reach better and faster results than the ones obtained through the slow comparative study of different organisms.

It was at this stage that the idea of expanding research on the plant beyond its molecular level started to gain new ground in the community. Given that the goal of AGI had been successfully fulfilled, why not drastically increasing the ambitions of Arabidopsis research and, as a consequence, its status as a model organism? Already since the 1980s, Meyerowitz, the Somervilles and Koornneef, among others, had entertained the prospect of using Arabidopsis as the material reference point for integrating knowledge coming from different biological disciplines. Even the staunchest advocate of molecular research felt uncomfortable in limiting investigations to the molecular level – and the need to combine research about the physiology, ecology and evolutionary history of Arabidopsis with the available genetic knowledge became even stronger in the context of geneticists' growing interest in functional genomics. The realisation that genetics alone would not have yielded insight in the functioning of organisms as wholes brought the most prominent members of the Arabidopsis research community to value more and more the plant as an 'intact'[66] source for knowledge about organisms as complex wholes. In other words, Arabidopsis was not interesting only as a representation of plant genes and pathways. More ambitiously, it could be taken to represent any whole organism as an ecologically adapted product of evolution: Arabidopsis had the potential to become an epistemically powerful tool for improving scientific understanding of the connections among processes occurring at different levels of organisation of the organism (genetic, cellular, physiological and morphological).

I should note here that the special representational value bestowed on Arabidopsis in the last few years does not in any way imply a *belief*, on the side of the scientists making that commitment, that one plant can indeed be taken as a representative of the whole floral kingdom. The current epistemic status of Arabidopsis specimens is, rather, dictated by the practical necessity of focusing research efforts on one organism. In fact, Arabidopsis researchers – not to mention biologists working on other model organisms – do not hesitate in indicating the many phenomena that Arabidopsis cannot be representative for. These include, for instance, RNA interference (injecting RNA into an embryo to see what happens), which can be done on Drosophila but not on Arabidopsis[67]; homologous recombination (which is better done on yeast); and several studies of plant pathogens, which would require an organism with a much longer life span than Arabidopsis. Further, there are of course many features that Arabidopsis cannot be representative for because it

[66] The following quote from Rubin, originally meant as a comment on the usefulness of Drosophila within animal research, well illustrates the reasons for favouring an intact organism: 'Many problems in eukaryotic cell biology can be most easily studied in unicellular organisms, such as yeast, or in cell cultures derived from multicellular organisms. Other problems, however, currently can be studied meaningfully only in intact animals. This may be because we do not know how to mimic crucial aspects of the organismal environment in vitro, because cell-cell interactions play an important role, or because the process under study involves a behaviour that is not currently understood in terms of the properties of individual cells. Examples include pattern formation in the embryo and the development and function of organ systems, such as the nervous system' (Rubin 1988, 1453).
[67] The reasons for this are unknown to biologists, one of whom remarked to me that 'it's meaningless to ask why, actually, it's like asking why it is so easy to transform Arabidopsis'.

does not possess them – like fruit development (for which tomato is a favourite candidate), nodulation (a symbiotic relationship with bacteria that allows legumes to fix nitrogen), adaptations to extreme climates (desert or seaside) and studies of chromosome shape (for which Arabidopsis chromosomes are too small). The awareness of these shortcomings does not, however, constitute a reason to challenge the current epistemic status of Arabidopsis – simply because, as most researchers argue, there is still so much to be understood and investigated about this plant and it is important to keep pursuing the centralised, well-coordinated research effort that holds it as its focus.

Here is how a 2004 report issued by the MASC re-phrases the goals of using Arabidopsis: 'the intent is that the knowledge gained on this experimental model organism will serve as the central reference and conceptual framework for all of plant biology' (The Multinational Coordinated *Arabidopsis thaliana* Functional Genomics Project 2004, 7). The authors go on to specify that 'the ambitious goal of understanding the function of all Arabidopsis genes as a first step toward an in-depth understanding of the biology of higher plants to the benefit of our society can be achieved, if sufficient and sustainable research funding is secured, biological materials and services are made available around the world, and human resources are further developed'. This new series of commitments is exemplified by the creation of two extensive projects officially initiated in 1999. The first is Project 2010, a ten-year-long umbrella project for the development of Arabidopsis functional genomics. In a nutshell, its goal is 'knowing the function of all plant genes by the year 2010' (Somerville and Dangl 2000, 2077). The second is The Arabidopsis Information Resource [TAIR], an internet portal aimed at gathering, storing and providing access to as much Arabidopsis data produced about Arabidopsis as possible (Garcia-Hernandez et al, 2002).

It is to this second project, and to the subgroup of the Arabidopsis community that is in charge of carrying it on, that I shall turn my attention. Originally created as an expansion of the AGI project[68] and a complement to existing Arabidopsis databases such as MIPS in Munich and NASC in Nottingham, the goals of TAIR actually far surpassed the already gargantuan tasks of storing and organising data at the genomic level.[69] As I shall argue in Chapters 5 and 6, the construction of dabatases to be included in TAIR involves not only the collection of Arabidopsis data from multiple sources (publications and reports by laboratories from a variety of disciplines), but also the elaboration of modeling strategies through which this ocean of data can be organised, retrieved and visualised by TAIR users. TAIR thus becomes a virtual laboratory experimenting with different ways of representing biological phenomena, such as Arabidopsis metabolism and biochemical

[68] In view of the huge amount of data expected to stream out of Project 2010, TAIR was initially thought of as a necessary expansion of Arabidopsis the Database (AtDB). Based at Harvard, AtDB was in turn the child of an initial database called AAtDB and it had the function of storing data coming from AGI, which it would do in collaboration with two other major European databases, the MIPS in Munich and the NASC database in Nottingham.

[69] Indeed, the epistemic goals proposed by TAIR are judged to be somewhat too pretentious and theory-informed (i.e. losing attention to the actual empirical data) by researchers working at MIPS and NASC. On the other hand, these characteristics are precisely what make TAIR research so appealing to a philosopher's eye. I shall say more about the difference between TAIR and NASC modeling strategies in chapter 5.

pathways. Most importantly, TAIR representations of these phenomena are taken to hold for the whole of plant biology. In the words of TAIR Director Sue Rhee,

> Ultimately, our goal is to provide the common vocabulary, visualisation tools, and information retrieval mechanisms that permit integration of all knowledge about Arabidopsis into a seamless whole that can be queried from any perspective. Of equal importance for plant biologists, the ideal TAIR will permit a user to use information about one organism to develop hypotheses about less well-studied organisms (Rhee website, accessed January 2005).

For my current purposes, two features of this mission statement are particularly remarkable: the emphasis on Arabidopsis knowledge as a 'seamless whole' and the high representational value placed on Arabidopsis as a model for other, 'less well-studied' organisms. This rhetoric, which pervades much of the programmatic descriptions of TAIR as well as acting as heuristic guide in its actual development, indicates precisely the shift in epistemic value of Arabidopsis as a whole organism. Massimo Pigliucci, one of the initiators of the efforts to integrate the ecological study of Arabidopsis with the analysis of its genetics, talks about this view as fitting a specific interpretation of the controversial notion of 'integrative biology': that is, 'a quest toward a comprehension of the connections among levels of biological hierarchy', rather than toward the formulation of a unifying biological theory or framework (Pigliucci 2003, 304). A first step towards such quest could be, according to many Arabidopsis researchers, a common focus and coordinated investigation of a unique organism.

The importance of TAIR with respect to the goals of Arabidopsis research as a whole, and especially to the pursuit of integrative biology, makes it an extremely interesting locus for analysing how Arabidopsis research might increase scientists' understanding of plant biology. Given that a close analysis of the thousands Arabidopsis-related projects would be impossible, I thus selected TAIR as one of two subgroups in the Arabidopsis community whose practices I want to explore in detail (I shall introduce and discuss the second subgroup in the next section).

It might be surprising to think of a bioinformatic project, a website, as a powerful stage for the ambitious pursuit of integrative Arabidopsis biology. A brief look at the state of contemporary bioinformatics will, however, quickly dispel the philosophers' scepticism. TAIR is part of a huge research effort – valuable both as a contribution to bioinformatics and as a development in theoretical biology – to find pragmatic and conceptual tools in order to integrate data, theories and models coming from different disciplinary backgrounds.[70] Bioinformatics is at the moment the most innovative and progressive field for data collection and integration in biology, generating a wide range of software, employment opportunities[71] and knowledge. In its integrative mode, Arabidopsis research

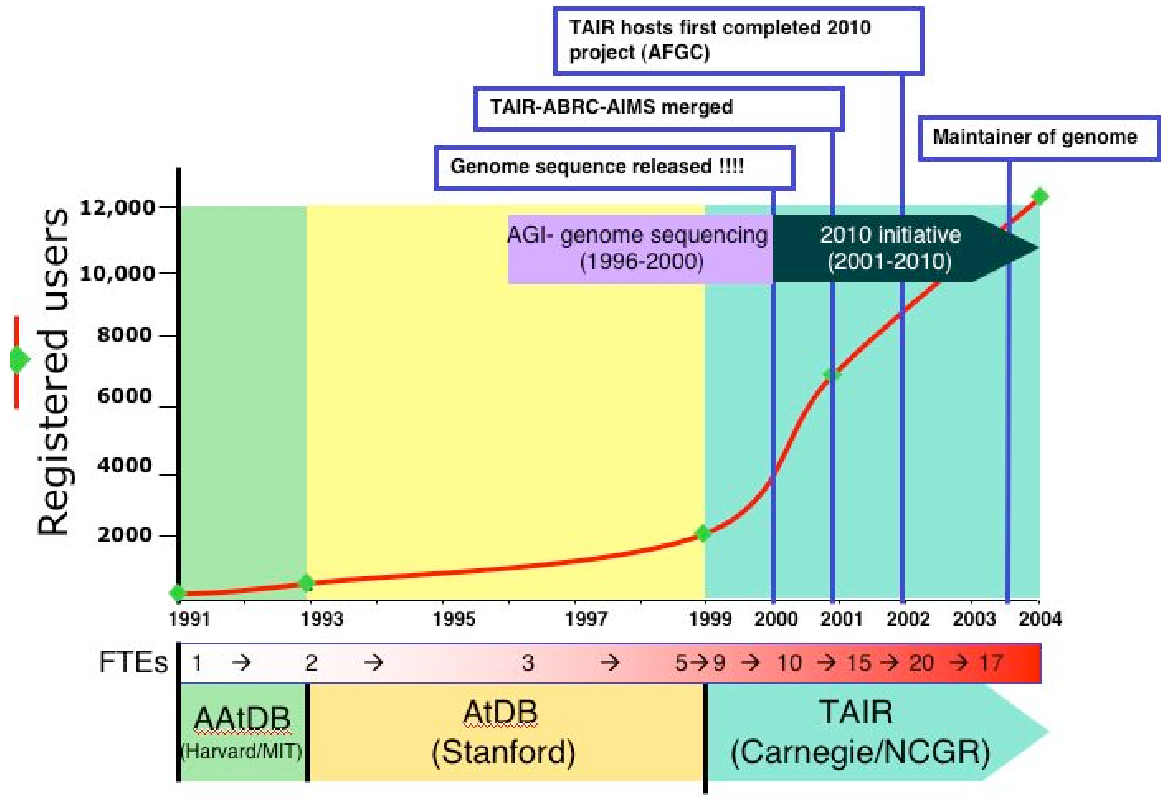
[70] See the enthusiastic 'update on bioinformatics' published in 'Plant Physiology' (one of the prominent journals in plant biology) by Somerville et al. already in 1997.
[71] The 'bioinformatician' is now a popular and sought-for interdisciplinary specialist, partially as a result of the needs of model organism communities (see for instance the recommendations for the enhancement of training in this field by the National Plant Genome Initiative, 2002).

plays a major role both as a propeller of IT innovation and as a beneficiary of new computational tools. Conversely, a close scrutiny of bioinformatics and of the tools that it can and cannot provide to this aim is crucial to understanding the modes of integration proposed to the Arabidopsis community. As is reported in 'Nature', 'the key to bioinformatics is integration, integration, integration' (Chicurel 2002, 751). Biologists and funding agencies (also in the EU, as testified by the large share of funding recently devolved to it) seem to agree that the key to data integration is indeed bioinformatics, insofar as facilitating data collection, exchange, elaboration and visualisation (in some cases, notably the newest 'workflows', by providing interactive tools that forge automatic connections between a variety of databases and elaborating results; in other cases, such as GenomeBench, by predicting genetic structures on the bases of previous data and even designing ways of testing such predictions). This attitude is mirrored in the NSF-funded Plant Genome Initiative that is currently backing Arabidopsis research.

The central role played by TAIR in the Arabidopsis community is evidently part of this trend. Already in 2000, at the end of the AGI project, TAIR had become the most accessible and recognised platform for exchanging resources about the biology of Arabidopsis. The very existence of the TAIR, currently counting more than 12.000 registered users (building up to more than 30.000 hits by unique IP addresses per month), underscores the existence of a community of researchers increasingly trying to construct a common terminology and a set of common goals for their multi-faceted research programmes. Sue Rhee's position as curator of the TAIR is crucial, since her team is in the process of considering and selecting the bioinformatic tools that provide the most suitable platform for collaboration among disciplines using entirely different types of data (e.g. gene sequences and images of morphological development of flower organs). Her work embodies one of the major challenges in contemporary bioinformatics: the construction of additional layers of data annotation which might fill the gap currently felt to exist between gene sequencing (resulting from the AGI and the HGP, among others) and the biology of organisms.[72]

Figure 3.5 - History of Arabidopsis Databases, starting from AAtDB and ending with TAIR, with the red line indicating the swift growth of users following the shift to TAIR.

[72] In the long run, innovative ways of annotating and elaborating data should enable a better understanding of the mechanisms governing what is known as the 'epigenetic space'. In turn, this is likely to result in a better overall understanding of the relation between genotype and phenotype - an advancement that according to most biologists would amount to understanding the organism as a whole, thus fulfilling one of the main research goals of the Arabidopsis community. Initiatives such as BioMOBY, an open source produced by the model organism community for biological web services, are also prominent in the attempt to make biological data increasingly accessible, usable and free from commercial obligations. According to its curator BioMOBY, the project 'was established to address the problem of discovering and retrieving related pieces of biological data from multiple hosts and services by attempting to generate a standardised query and retrieval interface using consensus object models'. The system 'facilitates the 'wandering' through large sets of data sets in a manner similar to the thought process biologists use when approaching their problems' (Wilkinson and Links 2002).



A short history of Arabidopsis Databases

Apart from the construction of the databases, the TAIR team is entirely committed to actively encourage interdisciplinary interaction among the various biological disciplines investigating Arabidopsis, to the investigation of new tools for integration (most notably in the form of bioinformatics and 'systems approach' experimental methodologies) and to creating educational tools targeted both to researchers and to interested laypersons. I shall delve in the details of TAIR integrative project in chapter 5. Let us now go back to the broader picture, that is, the decision by the Arabidopsis steering committee to re-focus Arabidopsis research towards integration. Admittedly, this represented a large step from the previous use of Arabidopsis as a model of plant genetics. First of all, the plant was assumed to be representative not only of flowering plants, but indeed of any plant. Second, it was not only the components of a plant specimen, such as its chromosomes and flowering system, that interested researchers: the plant as a whole was seen as valuable, thus coming to represent the intact organism whose complexity and cohesion keep mystifying all branches of the life sciences.

This highly pretentious goal was then (and still is) met with wide resistance. This is understandable, especially in view of the clash in commitments, research goals and strategies characterising biologists working at the micro-structural level with respect to biologists investigating ecological and evolutionary processes. The widespread resilience to attempts to integrate these types of research derives, among other things, precisely from the implied commitment to focusing research efforts on a limited number of organisms – an approach that strongly collides with the emphasis on diversity and comparative studies characterising, for instance, ecological research. It then comes as no surprise that Pigliucci himself found his closest peers to be hostile to his willingness to capitalise on the extensive molecular knowledge accumulated on Arabidopsis. Interestingly, a common objection was that the weed 'wasn't a real plant' (Pennisi, 2000), in the sense that it does not possess several characteristics deemed of crucial importance for the understanding of plants – such as, most blatantly, the capacity to bear fruits. This objection encapsulates both the above-mentioned perception of Arabidopsis as a model system for plant genetics, rather than a model of plants as whole organisms, and the belief that the actual plant specimen had little (if any) significance in the context of research at the molecular level.

But was it really the case that specimens had no significance within Arabidopsis research? A strong argument against this view is provided by the central role played by the centres for the production and distribution of Arabidopsis plants, starting precisely from the early 1990s. I already mentioned the importance of standardisation procedures in collecting and distributing specimens with relevance to the establishment of Arabidopsis as a botanical Drosophila. As it turns out, the institutionalisation of techniques, resources and categories with which to grow laboratory specimens played an even more fundamental role in the most recent and ambitious phase of Arabidopsis research. Accordingly, I shall now introduce the Nottingham Arabidopsis Stock Centre, which constitutes the second locus of Arabidopsis research practice on which I focus my investigation.

3.3.2 Growing models: the Nottingham Arabidopsis Stock Centre [NASC]

The shift in the representational value of Arabidopsis plants implied the recruitment of researchers from a variety of fields to work on a model organism hitherto envisaged as a model system for plant genetics. The expansion of goals had one further important consequence: a strengthened need to make Arabidopsis specimens easily available to biologists interested in working with them. At the practical level, it seems easy enough to fulfil this requirement: all that is needed is the creation of structures whose responsibility is to provide biologists with the specimens that they require. Yet, given the great variability among Arabidopsis ecotypes and the proliferation of mutants made possible by the Feldmann technique, this seemingly simple task becomes exceedingly arduous. Let me now clarify some of the practical hurdles associated with such a massive standardisation project. For a start, criteria need to be in place for the description and categorisation of different Arabidopsis specimens, which would be understood and used by the whole community independently of the expertise of each participant subgroup. Also, stocking and distributing Arabidopsis specimens requires a high level of control of the quality and genetic make-up of the seed. The risk of contamination among different lines, for instance, is enormous. In a typical laboratory glasshouse, a slight current is enough to transport some of the thousand microscopic seeds produced by a single Arabidopsis plant across a room and into a different harvest, thus causing seeds of different type to mix. In sum, the production, storage and distribution of different types of Arabidopsis specimens is as difficult a task as it is indispensable to the Arabidopsis community. It requires both technical and scientific expertise; constant interaction with the community as a whole concerning the latest experimental practices, techniques, instruments and results; and the capacity to interpret research results so as to make informed choices about which are the most promising fields in Arabidopsis research (so as to be able to produce and supply enough of the relevant specimens to fit requests).

Most importantly for my purposes, all these activities have a great influence on scientific reasoning via model organisms as well as on the understanding derived from model organism research, since they affect the experimental procedures used on Arabidopsis specimens, the interpretation of such procedures as well as the very nature of the specimens themselves. This is because Arabidopsis scientists, no matter their specific expertise or location, need to be sure that their experimental work is based on relatively similar plants. All the efforts accompanying the decision to focus research on one specific organism, thus collecting data concerning different aspects of the same organism and holding that 'integrated understanding' as representative for other plants, would be in vain if it turned out that researchers had been investigating Arabidopsis ecotypes and mutants with very different characteristics – which could therefore hardly be defined as 'the same organism'. In order to ensure that theoretical integration among Arabidopsis findings can take place, variation among individual plants has to be eliminated via a highly controlled and centralised process of standardisation. Researchers need to be sure that they are gathering results on organisms possessing the same material features (natural and/or induced), in order to then compare those results and compile them into a unique dataset about that organism. The indiscriminate use and classification of Arabidopsis variants cannot be tolerated, if results are to be replicable, extended and/or used in several research sites (each of which cultivates its own plant collection). Finding a solution for this problem constituted as important a step towards pursuing the ideal of integration as the construction of digital data repositories such as TAIR. Of course, the standardisation of plants does not constitute integration as such: yet, it is a necessary material platform towards obtaining such integration at a theoretical level.

The MASC itself is aware of how relevant the categorisation and production of specimens is to biological research practice and to the interpretation of its results. Already in the mid-1980s, the Steering Committee of MASC could foresee the extent of the difficulties and problems gravitating around specimen production. It therefore proposed to establish a single Arabidopsis stock centre, a central institution that would differentiate among lines and ecotypes, dictate standards and protocols for recognising and handling them, and generally assist researchers by supplying specimens with the required features. As a result of this proposal, the Nottingham Arabidopsis Stock Centre [NASC] was created in 1990. Up to that point, one of the main collections of Arabidopsis specimens had been maintained in Frankfurt by the successor of Röbbelen, Albert Kranz, who was planning to retire by the end of the 1980s. The MASC committee decided to take advantage of that change of guard to transfer the collection to a new facility in the Plant Science Institute in Nottingham, UK. In 1991, a twin facility was established at Ohio State University in order to facilitate distribution of seeds to American researchers, as well as to guarantee that there would be a duplicate of every strand stored by NASC (in the event of fire or other disaster affecting the storage rooms). The Arabidopsis

Biological Resource Centre [ABRC] and NASC were conceived as working in strict collaboration – and indeed, over the last two decades their two directors have been working closely together in order to create and implement common standards and procedures.[73] I shall devote the final paragraphs of this section to introducing NASC activities and their relation to Arabidopsis history as a whole, so as to provide a basis for the in-depth analysis of NASC practices and their epistemological significance that I will offer in Chapter 6.

NASC has much broader responsibilities than its German predecessor. For one thing, it has to store and categorise the thousands of knockout lines produced thanks to Feldman's Agrobacterium-induced mutations. As mentioned above, the rapid creation of so many genetically modified strands of Arabidopsis specimens requires a much stricter policing of the lines, as well as new methods for handling the storage and maintenance of that many plants. Standardisation and the related control acquired over lines guarantees the uniformity and reproducibility of results, by allowing scientists working at opposite ends of the world to credibly claim to be working on the same plants, grown in similar conditions and from the same stock.[74] For instance, Feldman's knockout lines were all produced with Wassilewskija [Ws] as a background line, an ecotype that he regarded (wrongly, as it turned out) to be more easily transformed than its more popular relatives Col and Lan. The perceived necessity to replicate Feldman's experiments and pursue his line of research on the same plants that he had used enormously enhanced requests for the original Ws ecotype and mutated lines. In 1995, Feldman thus donated this whole stock, in the form of plant seeds, to NASC. NASC handled the donation according to its standard procedure. It dehydrated and froze a sample of the donation, so that it would be safely stored in its seed archive (which NASC members half-jokingly refer to as 'seed museum'). The rest of the seeds were then re-bulked under conditions of maximal separation from other lines (to avoid cross-contamination). The resulting stocks were then sorted, bar-coded, packaged and transferred to a low temperature, low humidity storage room, where they wait to be distributed according to demand from laboratories around the world. The resulting availability of the Feldman stock through NASC quickly transformed Ws into the third most popular ecotype in the whole community.

The shift in epistemic goals within the community did revive the interest in collecting natural variants (Alonso-Blanco and Koornneef 2000), since 'different ecotypes apparently do things in quite a different way' (Pigliucci 1998, 301) – thus making the exchange of stocks akin to a 'research currency' in the community, where the stock centres function in much the same way as the stock exchange.[75] Yet, while such wildtypes were used to target questions relating to ecological variation and adaptability, Col, Lan and Ws established themselves as the main tools for exploring epigenetic and cellular mechanisms for gene expressions as underlying development and growth patterns of plants. As a result of these and several further developments, NASC and ABRC had to

[73] A joint statement by NASC and ABRC directors, discussing their tasks and their visioen for future stock storage and distribution, can be found in Scholl et al (2000).
[74] For in-depth analyses of the role of standards as coordinating devices, see the seminal work by Bowker and Star (1999) and Berg (2004).
[75] I owe inspiration for this metaphor to Maarten Koornneef.

channel huge efforts into the standardisation of seed storage and distribution: new ways to distinguish wildtypes from their mutated offspring; tools and techniques to avoid cross-contamination among strands (given especially the fact that they had to be grown and harvested in the same limited space); and methods and space for storing multiple copies of both wild and mutant lines.

Given the difficulties encountered by stock centres themselves in growing the plants, and thus their increasing expertise in handling them, it became their responsibility also to transmit these skills to their 'users' (that is, the researchers requiring and using the seeds for their own research purposes).[76] This meant issuing strict guidelines concerning the conditions in which plants should be sown, germinated, grown and harvested (temperature, humidity, type of soil and pesticides, etc.).[77] The ensemble of protocols, guidelines and actual modification of specimens elaborated by NASC and ABRC thus represents a notable amount of standardisation, given especially the ease of maintenance characterising Arabidopsis wildtypes (which after all, in their quality of weed, have the infamous tendency to grow pretty much everywhere). This demonstrates how, once adopted in the laboratory, Arabidopsis lines ('ecotypes' or 'accessions' according to different nomenclatures adopted by the stock centres[78]) became something different than their relatives living in the wild. They have to be kept in the best possible growth conditions; to be easily distinguished on the basis of standard morphological parameters, such as the width and smoothness of the leaves, the shape of the flowers and the length of the stem; and to bear the material features most compatible to the research purposes for which they were used. Such modifications of the 'wild' specimens are partly determined by the need to store, produce and distribute the plants on a massive scale.[79] At the same time, most changes are effected to 'serve the user' (in NASC terminology), i.e. to fit the practical needs and epistemic goals of researchers using the plants. Either way, the plant features thus obtained can all be classified as induced characteristics, according to my categorisation in 3.2.3.

[76] Further, stock centres have to handle a good deal of PR, since they have to actively encourage researchers to donate samples of the plants that they used for each experiment. In line with the Arabidopsis ethos, these donations are expected to come free of charge and as a service to the community – yet, as testified to me by the current director of NASC, Sean May, eliciting donations is no easy task even within the collaborative environment of Arabidopsis research (as exemplified by the need for Material Transfer Agreements for some of the mutant lines provided by NASC).

[77] ABRC guidelines are 'designed to generate healthy plants that give maximum set of pure seeds and to preserve these in the safest and most convenient manner' (ABRC website, accessed 6 May 2005).

[78] Throughout this chapter, I have been employing the terms 'ecotype', 'accession', 'strand' and 'line' as synonymous. While acceptable for my purposes, this usage effectively eliminates a major source of disagreements within the Arabidopsis community. In particular, ecologists and molecular biologists are divided over the meaning of the term 'ecotype'. Most molecular biologists take the term to refer to any accession (i.e. a group of plants with the same genotypic structure) derived from a wildtype (i.e. an organism not genetically modified). Ecologists, on the other hand, would only talk about ecotypes as indicating the phenotypic traits of an organism that result from genetic adaptation to its environment. Unfortunately, given the potential interest of an analysis of these discussions, I cannot focus on these issues within this dissertation and shall therefore keep using these terms as synonymous.

[79] This is well illustrated by the zwapak example that I provide in Chapter 6, section 6.2.1.

3.4 Understanding via Model Organisms: Three Areas of Research

In this chapter, I have presented the history of biological research on the model organism Arabidopsis thaliana, with specific focus on two current projects within the Arabidopsis community whose main purpose is to produce tools enabling an integrative understanding of plant biology. In section 3.2, I sketched a categorisation of the features of Arabidopsis that, throughout the last three decades of research, most contributed to the growth of biological knowledge about the plants and, hence, to the success of Arabidopsis as a model organism. In this sense, these features can be defined as epistemologically advantageous to the acquisition of scientific knowledge. However, I have not yet explained how this is the case. Nor have I discussed how working on a model organism that possesses these characteristics helps biologists' *understanding* of plants. The study of how Arabidopsis came to be known as prominent tool towards scientific understanding constitutes, in fact, a platform from which I can start tackling these crucial questions.

Consider again the list the epistemologically advantageous characteristics of Arabidopsis that I provided in section 3.2. On the basis of that list, I shall now distinguish three main aspects of model organism research, each of which I take to signal a fundamental condition towards acquiring a scientific understanding of organisms. The first aspect concerns the *embodied knowledge* needed to deal with the material features of organisms: that is, the constraints imposed on researcher's actions and thoughts by the physical nature of some features of Arabidopsis specimens and thus by the need to learn how to intervene effectively on them for research purposes. Of course, biologists manipulate the material features of an organism in order to acquire knowledge about its functioning and its structures. In other words, they use their embodied knowledge of model organisms to acquire *theoretical knowledge* of their functioning and structure. This aspect is especially evident when considering the projected characteristics of Arabidopsis (point 3). Without even starting to examine the content and quality of scientific results obtained through the study of a model organism, these features illustrate the extent to which Arabidopsis research is useful as a thinking tool and heuristic guide to researchers. The last three elements on my list of Arabidopsis features, which I call 'characteristics of the community', define the co-dependence characterising Arabidopsis specimens and the community within which such specimens are studied. On one hand, the currently thriving Arabidopsis community owns its interdisciplinary and multi-sited character to the decision of focusing efforts on this plant. Arabidopsis plants can be argued to constitute 'boundary objects' in the sense outlined by Star and Griesemer: material representations of nature that bring together social worlds that are usually separate from each other.[80] On the other hand, Arabidopsis plants would not have the status of model organism (and thus would not be produced, distributed and used as such) unless they had been adopted and promoted as such by a community of capable biologists. Thus, the *social organisation* of communities conducting model organism research constitutes a third crucial aspect.

[80] Perhaps surprisingly, there are many analogies between the case analysed by Star and Griesemer (that is, natural history collections in musea) and the case of model organisms. See for instance their description of the pre-conditions for the successful establishment and use of boundary objects within a community (1989, 408).

These three aspects of model organism use in biological research constitute the focus of the rest of this dissertation. The relation between embodied and theoretical knowledge of the model organism Arabidopsis, as used by biologists in the whole Arabidopsis community, will be analysed in Chapters 5 and 6. Chapter 7 will focus on the crucial role of social epistemology in enabling the understanding of models and theories about biological phenomena that are gathered via the study of Arabidopsis. Via a detailed study of the practices and skills characterising the Arabidopsis community, I shall identify and study some of the conditions under which a biological understanding of plant biology is gathered in the community (Chapter 8). Before venturing into my analysis of model organism research, however, I wish to voice some reflections on the methodological issues emerging from research that, as I propose to do here, deals with philosophical, historical and social aspects of biology at the same time. Chapter 4 is thus devoted to presenting and defending the research method that I refer to as ‘complementary science’.

Chapter 4. Methodological Framework: Is Empirical Philosophy a Philosophy of the Particular?

Understanding is not achieved by generality alone, but by a relation between the general and the particular
Richard Levins 1984, 26

The three scholarly traditions of philosophy, history and sociology of science have increasingly diverged over the last few decades. Despite the examples set by Bruno Latour, Peter Galison and James Griesemer, among others, attempts to integrate 'science and technology studies' [STS] with 'history and philosophy of science' [HPS] remain rare and are often the object of controversy.[81] One of the charges levelled by some philosophers to the work of historians and sociologists concerns the limited scope and applicability of their results. History and sociology, it is claimed, study the local, the situated, the contextual; their narrative is unavoidably a narrative of the particular, given the richness and relevance of details that these empirical disciplines need to analyse. Philosophy of science, on the other hand, is a largely rationalistic enterprise whose aim is to reflect on the general, the abstract, the universally applicable. It uses historical examples as exemplifying some or other conceptualisation, yet it should not let itself be deranged by reference to a handful of cases: its aim is to produce ideas that are not tied to a single occurrence, but that can inspire and be used in several different contexts and situations. Empirical philosophy of science, according to this type of rationalism, is destined to generate ideas that are extremely limited in their applicability, since their scope is constrained to the analysis of a handful of practices. A philosophy of scientific practice, in other words, can be nothing else but a philosophy of the particular, with nothing to offer to philosophy of science as a whole, let alone to scientists or other science scholars.

Chapter 2 already contained a philosophical critique of this strict form of rationalism, as represented in the work of Carl Hempel. This lead to a defence of an empirical philosophy focused on the specificities of scientific practice. In this chapter, I intend to focus on the *how* questions, that is on the methodological issues that arise when actually pursuing a philosophy of scientific practice. In particular, I consider the difficulties

[81] Both the philosophical and the STS community have contributed to widening the divide between the two disciplines. Influential philosophers such as Larry Laudan (1990) and Philip Kitcher (1993) have warned their colleagues against meddling with sociology. At the same time, some STS scholars have stirred younger generations of STS students away from philosophy and towards more 'applied' and 'concrete' types of analysis. The dismissal of philosophy as part of STS is visible from the list of 'core literature' adopted by the Dutch society for STS research (WTMC), which only includes four outdated philosophy references (Popper 1963, Kuhn 1970 and two old reviews by Hans Achterhuis) over 66 social scientific volumes (http://www.wtmc.net/wiki/index.php?title=WTMC_list_of_core_literature, accessed 24 July 2006). In addition, disciplinary boundaries are enforced at the institutional level, for instance by the establishment of strongly field-centred journals that refuse to publish pieces 'contaminated' by other disciplines. Within philosophy, high-profile journals such as 'Philosophy of Science' and the 'British Journal for the Philosophy of Science' are reluctant to publish research informed by empirical studies on specific contexts. In a similar way, history and sociology journals such as 'Annals of Science' and 'Social Studies of Science' have very high standards on the quality of the evidence submitted and generally do not accept largely theoretical analyses.

encountered when reconciling a detailed study of cases with the production of general claims. The present piece of research is a good exemplification of the seemingly paradoxical fusion of particular and general that haunts empirical philosophy. On the one hand, I tackle one of the most widely applicable, all-encompassing and significant topics in science and technology. My philosophical study of the conditions under which scientific understanding is achieved involves a reconstruction of what I take to be relevant components of scientific practice (and indeed, I propose such a reconstruction in Chapter 8). The resulting epistemological analysis could, in principle, be applied to all biological science, or even to science as a whole. It aims at generating discussion and accordingly, in the absence of further research, no constraints are imposed a priori on the applicability of my conclusions. On the other hand, the range of scientific practices on which these results are based is as limited as they are inclusive. I am considering a single case study, the Arabidopsis community, which, despite its size and internal diversity, is hardly representative of all research in the life sciences. Moreover, a detailed study of the practices characterising contemporary biology requires methodological input from the history and sociology of science. Reading practitioners' own accounts of their activities is not enough, when it is also possible to witness them in person. The need to visit laboratories and gather scientists' private opinions about what they are doing forced me to borrow methodological insights and procedures from both HPS and STS, so as to be able to extract arguments and claims from the large array of data gathered by participating in, rather than simply observing, Arabidopsis biology.

This chapter reflects on the methodological problems associated to conducting the mixture of sociological, philosophical and historical research that is ungracefully subsumed under the heading of 'history, philosophy and social studies of science and technology' [HPSSST]. My discussion assumes that an interdisciplinary approach to the study of science, such as the one attempted in this thesis, is both useful and feasible[82]: it therefore explores the nature and epistemological significance of the practical dilemmas encountered with carrying out this type of research. These reflections are fuelled by the experiences that I accumulated while conducting the present inquiry. A secondary aim of this chapter is thus to illustrate and account for my choice of materials, the way in which I gathered them and the way in which I used them to feed my epistemological pronouncements.

4.1 Studying Science and Technology: Insularity and Integration

I begin by dwelling briefly on the current state of interdisciplinary approaches to the study of science. What causes the lack of interaction among scholars of science that

[82] As I specify below, I do not intend to argue that such an interdisciplinary approach is *always necessary* to the study of science (regardless of the issues under investigation); nor do I want to claim that integration among the different disciplines involved in the study of science should be conducted in one and only way. As Wim van der Steen has convincingly shown, an excessive and uncritical emphasis on the integration of disciplines constitutes 'a modern variant of the old ideal of unity of science' (1993, 259), especially when interdisciplinary methodology is seen as one method with standard characteristics and universal applicability. I do not think that this argument applies to cases where the manner and purpose of integrating different approaches is motivated by and adapted to the actual issues under investigation.

HPSSST attempts to challenge? Besides the structural and social constraints dictated by specialised training and disciplinary expertise, is there anyone to blame for the poor integration among philosophical, historical and sociological approaches? And are there exceptions to this trend? Discussing these questions will hopefully set the background against which to examine problems and methods associated to HPSSST research. In what follows, I illustrate the relevance of an integrated approach to the study of science by pointing to some shortcomings of what I call ‘insular’ philosophical, historical and sociological research.

Before setting out, it is important to note that this is not a manifesto requiring every scholar of science to contribute to all three disciplines at once. As I argued elsewhere, interdisciplinarity is not the achievement of individual scholars. It is primarily the result of cooperation among researchers of varying expertises, gathering together in order to confront a common problem (this is the so-called *issue-based* mode of research; Leonelli, 2004 and 2005).[83] A general critique of disciplinary training as such is therefore far from my present purposes. Further, not all issues emerging from science and technology need to be investigated at all times by an interdisciplinary research group. For instance, the nature of laws and concepts in science is a largely philosophical concern, while the organisation of research communities and their relation to funding sources are matters of inquiry for a sociologist. I have no qualms with exploring such issues through firmly discipline-centred methods (which have in fact been revised and improved throughout decades of trials and errors for precisely this purpose). What I intend to condemn is the inability, currently displayed by a majority of scholars of science, to learn from other disciplines’ results and to form interdisciplinary teams when required by the nature of the topic at hand. Interdisciplinary discourse across philosophers, sociologists and historians is still very limited. It is difficult to organise research venues and launch publications focusing on intersections among expertises. Also, the abundance of literature in one’s chosen field discourages from venturing into other disciplines to check for relevant knowledge. This insularity represents a loss to all research communities involved. This is certainly the case with phenomena like biotechnological innovation or nuclear energy, whose examination requires interdisciplinary brainstorming. A lack of exchange also affects the development of areas that might not require interdisciplinary research, but might nevertheless benefit from insights from different areas. For instance, while it is true that the nature of concepts is an issue of primary interest to philosophers alone, the consequences of such analysis are relevant to both historical and sociological uses of concepts (as demonstrated by the recent come-back of the Koselleck-inspired field of ‘conceptual history’ as an extremely fruitful historical approach). At the same time, the scientific application of concept-based approaches to actual cases could teach much to the philosopher willing to lend a hear. Similarly, as I show in this dissertation, the

[83] I am aware of the literature documenting the difference between interdisciplinarity and transdisciplinarity, where the former characterises collaboration across disciplines and the latter involves collaboration on a specific topic. For the purposes of this thesis, I shall not take this distinction into account. This is because, in the context of my analysis, it is not helpful to separate issue-based research from research requiring dialogue across epistemic cultures (in fact, it may prove entirely misguided, as interdisciplinary collaboration always requires ‘transdisciplinary’ attention to common topics or objects).

sociologist’s take on community structure and funding has serious implications for an epistemologist’s analysis of models and explanations produced by a research community.

Let us consider this last point more closely. Very few *philosophers* trained in the Anglo-American tradition are interested in bringing sociological and historical insights into their analysis. This is true also of scholars that rejected old-school rationalism and concern themselves with current scientific practices, in accordance to the recent empirical turn that I discussed in Chapter 2.[84] Given their reluctance towards adopting interdisciplinary insights, these philosophers might be portrayed as prime culprits for insular discourse. Helen Longino has attributed the prolonged philosophical disinterest in social aspects of science to what she calls the ‘rational-social dichotomy’ (Longino 2001,1): namely, the pervasive notion that the nature of science can be defined as *either* rational *or* social – and that focusing on one of these two aspects automatically excludes consideration of the other. This influential idea brings many philosophers to disdain the work of sociologists, as in their eyes attention to the social aspects of science (which they see as synonymous with ‘contextual’) corresponds to a denial of its rationality, truth-value, empirical adequacy and so forth.

I agree with Longino that the rational-social dichotomy is entirely misguided. Nevertheless, philosophers are not the only scholars of science to have fallen under its spell. Outright opposition to the interest in rationality and truth expressed by philosophers was a key motivation towards the construction of one the most influential schools in sociology of scientific knowledge, that is the ‘strong programme’ put forward by Barry Barnes and David Bloor at the University of Edinburgh (Barnes and Bloor, 1982). While the strong programme presented several insights that would have benefited both the philosophy and the sociology of science, such as the ideas of symmetry and reflexivity, its proponents’ insistence on the opposition between the aims of sociologists and philosophers of science contributed to deepening the burrow already dividing the two camps.[85]

Sociologists busy with laboratory studies and ethnographic approaches were influenced by this anti-philosophical attitude, as made clear by Latour and Woolgar in the introduction to their topical ‘Laboratory Life’ (1979). Mistrust of the ‘rationalism’ and ‘black-boxing’ attributed to philosophical analysis led many sociologists to turn their backs to philosophers. Towards the mid-eighties, the increasing attention to the social values and interests involved in scientific practice also commanded a shift in the sociologists’ choice of objects of study. Research on the content of scientific knowledge as a social achievement declined dramatically, to be substituted by studies of the role of science as a social institution and of scientists as successful entrepreneurs, whose main abilities lie in gaining and maintaining recognition and social authority. The emerging

[84] This statement might sound paradoxical, as the empirical turn is supposed to imply precisely a renewed attention to historical and sociological research. However, many of the protagonists of this turn in philosophy (Giere 1999a, Kitcher 1993, van Fraassen 2002, Solomon 2001) have not actually engaged in empirical research on science practice: they have borrowed examples from studies of such practice and they have used them as an illustration of, rather than a potential challenge to, their own points.
[85] Sociologist David Bloor and philosopher John Worrall conducted a famous debate epitomising this controversy, which is documented in Worrall (1990).

need to understand recent advances such as the digital revolution, the advent of information and communication technologies and of other types of artefacts (such as nano and biotechnology) also contributed to this shift of focus. Technology has taken over science in the eyes of most sociologists, as demonstrated by the popularity of Wiebe Bijker's views on the social construction of technology (1995). Why spending time on the content of theories whose truth-value has been severely questioned by constructivist critiques, when processes such as the digital revolution are reshaping politics, social dynamics, trade and communication all around the world? This is an excellent and very reasonable question. Social studies of science have been extremely successful in highlighting the extent to which controversy and disagreements permeate scientific practice. Thanks to this work, we can no longer accept the idea that scientists unanimously agree on what constitutes scientific truths. This constitutes a great starting point for philosophical work on pluralism in theories as well as practices, which some philosophers have indeed welcomed and pursued. Many sociologists are not, however, aware of these latest developments. As a consequence, they keep portraying philosophy of science as an unquestioning defender of scientific rationalism and, at the same time, they pay much more attention to the social context in which research is carried out than to its scientific aims and results. Much more reflection is needed on the value and status of the objects (issues, concepts, situations, sites, concerns, policy measures, personal experiences or else) that STS investigates, as well as on the type of results that such investigations can and should yield.

The sociologists' disregard for the content and significance of scientific and technological knowledge constitutes the main difference between their work and the work of *historians* of science and technology. The latter group preserves a much stronger focus on the content, rather than just the context, of scientific knowledge. In order to comprehend how complex processes like theory-change and discovery take place, historians rightly insist on the necessity of descriptive and anti-Whiggish research (that is, on reconstructing the roots, significance and implications of such processes at the time of their occurrence[86]). Attention to actual scientific results notwithstanding, collaboration with philosophers proves difficult, given the characteristic resilience of historians towards systematic analysis and generalising conclusions. Historians are primarily interested in 'getting the facts right' – in other words, they see their main task as one of recovering knowledge that has been lost. This often implies deep reflection on what counts as a fact for the actors involved and its relation to the type of knowledge gathered and sanctioned by their community of peers. The fruits of such reflection are however tied to the specificities of the circumstances and events to which they are meant to apply. Rather than providing analytic tools for the study of episodes in the history of science, much historical research is directed at providing illustrations for specific themes and processes (such as, most recently, 'objectivity'[87]) and thus refuses to venture into the realm of theoretical analysis.

[86] See Nickles on 'reconstruction' in HPS (1986, 1995).

[87] On objectivity, see Daston (2000) and recent issue of 'Social Epistemology' by Boumans and Beaulieu (2004).

I take contemporary philosophy of biology to constitute a promising exception to the above-mentioned insular trends.[88] This is largely due to the history of this field, which has been marked in the 1970s and 80s by the charismatic influence of interdisciplinary contributors such as the philosophers-historians Marjorie Green and Everett Mendelssohn. It is also due to its peculiar institutionalisation, as exemplified by the International Society for the History, Philosophy and Social Studies of Biology (informally known as ISH Kabibble). Founded in 1990, this society strongly encourages its members to share insights ranging through all three disciplines mentioned in its title. In its quality of the one and only well-recognised international meeting point for philosophers of biology, the society has played a major role in shaping and directing research in the field through the last two decades. As a result, interdisciplinary research on biology is relatively safe from the disciplinary diatribes characterising other fields and receives substantial institutional support.[89] Accordingly, all prominent journals dedicated to the philosophy of biology favour a history-informed reconstruction and interpretation of biological practices over the more traditional focus on a priori reasoning that is still privileged in general philosophy of science.

History, philosophy and social studies of biology [HPSSB] thus provides a welcoming environment for my own research interests, that is for using methods characteristics of the history and sociology of science in order to achieve a philosophical analysis of scientific understanding. I shall now turn to examine the methodological implications of seeking to incorporate the concerns of both sociologists and historians – from the interest in social dynamics and technology to the attention to local, situated practices – without renouncing the philosophical ambitions to systematise, elucidate and critique the ‘facts’ of which science is made.

4.2 Reconsidering Research Goals: HPSSB as Complementary Science

Note that the above account of scholarly attitudes towards interdisciplinary research on science and technology emphasises the *intellectual* interests displayed by the philosophers, historians and sociologists involved. This choice is deliberate. While institutional and financial aspects are equally relevant to explaining the current state of affairs in science studies, the focus on research goals and values underlying commitment and/or resistance to interdisciplinary scholarship is more directly relevant to my methodological discussion. This is because I believe a critical examination of the *concerns* fuelling research to be tightly intertwined with the evaluation of which courses of action might appropriately serve those interests. A normative discussion of HPSSB methodology, as exemplified in my own work, needs to start from a reflection on the *goals* that research is supposed to fulfil. In this section, I shall therefore examine the goals and values governing my methodological strategies. As it will turn out in the next

[88] See Callebaut (2005).

[89] This claim is valid only at the international level and relatively to other fields in philosophy of science, since philosophers of biology have often trouble in making other faculty members of their local institutions, as well as local funding bodies, accept their integrating preferences. A brief analysis of this phenomenon in the case of the Netherlands can be found in Leonelli and Reydon (2005).

sections, a close analysis of how these goals might be achieved in practice does not lead to selecting one specific methodology as the absolute best HPSSB tool for inquiry. Rather, considering HPSSB 'in action' illustrates that this type of research is associated to a set of key concerns, rather than a set of methods. These concerns, including empirical adequacy, generality and reflexivity, are equally important to achieving HPSSB goals and should therefore be carefully calibrated in each specific piece of inquiry. The choice of methods that would support those concerns and facilitate results depends largely on the subjective preferences and context of each HPSSB researcher.

In order to describe the goals and values that I take to motivate my contribution to HPSSB, I draw inspiration from a recent programmatic statement made by philosopher and historian of science Hasok Chang. Chang defines his brand of HPS research with the label of 'complementary science', which he describes as seeking 'to generate scientific knowledge in places where science itself fails to do so' (Chang 2004, 236). Chang's starting point in building this position is the Kuhnian notion of 'normal science', which he takes not only to constitute the vast majority of scientific practice, but also to possess characteristics whose significance was crucially underestimated by Kuhn himself.[90] One of these characteristics is the tendency to close itself to the critical challenges that are necessary to its very development.[91] Imre Lakatos (1970) famously pointed to the necessity, in the context of increasingly specialised research programmes, of holding on to a 'research core' for guidance and constructive development. In an equally powerful pronouncement, Karl Popper condemned the dangers, to society and science alike, that lurk behind this partially dogmatic attitude (Popper 1970, 53). This debate generates crucial questions as to how to pursue the goals of normal science without suggesting drastic changes to its modes of operation (which, after all, have proved to be immensely efficient), but at the same time without being paralysed by the constraints on questions and commitments to which any research community, typically limited in its human and material resources, is inevitably subjected.

Chang proposes complementary science as an answer to this dilemma. Like science itself, complementary science aims primarily at a better understanding of *reality* itself, rather than of science as a system of knowledge. It is interested in 'normal' scientific research as a *subset* of that reality, whose results and practices are therefore subject to probing and scrutiny in much the same way as scientific phenomena are to normal scientists. Complementary science thus pursues the goal of understanding reality by examining science just as closely as the rest of the natural world on which scientists focus their attention.[92] Hence, its secondary goals, which mark its difference from normal scientific

[90] The argument that Kuhn's description of normal science, though insightful, does not go far enough in overseeing its implications, is certainly not new. The work of Steve Fuller (2000), Joseph Rouse (1987) and Peter Galison (1997), to name just a few, starts from the same claim and develops into very different, fascinating accounts of the nature of scientific practices.

[91] The importance of critical reflection within scientific community will be further emphasised when discussing Longino's social epistemology in Chapter 7, section 7.3.

[92] The assumption here is that scientists and their activities are certainly part of the animal (and thus the natural) world. It might be argued that the artefacts produced through those activities are not part of the same natural world, but it is not within the scope of this chapter to discuss this complex claim. I shall come

research, are the development of a critical perspective on the methods and aims of science, the search for alternative frameworks and methodologies and the proposal of new issues and agendas for the future.[93]

What type of research might fall under the heading of complementary science? Free from the restrictions of tight research goals, specialised questions and deadlines imposed by funding bodies, philosophy of science (and especially the philosophy of scientific practice) can indeed take upon itself the critical role that specialised science is often forced to forego.[94] The philosopher's work is, in this sense, overlapping with the historian's interest in uncovering and re-evaluating questions and ideas that have fallen out of the scientific agenda for all sorts of reasons. Further, and here I take issue with one of Chang's claims[95], contemporary history and sociology of science can also operate in such complementary manner. They can do this by addressing interests, circumstances and ideas that today's practicing scientists do not reflect upon or do not consider to be relevant to their research (as it often turns out that social, political and economic interests do have a great impact on both the contents and the directions of scientific and technological research). In this sense, HPSSB can serve as a type of complementary science. In fact, I maintain that at least some areas of HPSSB research, such as my own

---

back to it, in a somewhat different form, when discussing the difference between model organisms as artefacts and samples of nature in Chapter 6.

[93] In this sense, complementary science differs greatly from the naturalistic approach to scientific epistemology advocated, for instance, by Ron Giere (1988): a complementary scientist aims to critique and further the claims of normal scientists, while a naturalist wants to use scientific knowledge to produce philosophical arguments. I agree with Chang's Popperian stance that 'constructive scepticism can enhance the quality of knowledge, if not its quantity' (2004, 243): naturalist epistemology, however, leaves no space for such constructive scepticism.

[94] I do not mean to depict philosophers and historians of science as entirely free from professional demands, deadlines and various kinds of financial obligations. Whether they are located in scientific or humanistic departments or institutes, HPS scholars are certainly not immune from the pressure to compete with their peers, satisfy their sponsors and fulfil requirements of teaching and publication. Still, HPS professional constraints are relatively minor when confronted with the constraints and specialisation often affecting scientific research. For instance, the way in which I chose to develop my PhD research was entirely independent from funding or institutional matters. This determined financial difficulties in funding my field trips, which meant that I could not afford very long stays in Arabidopsis laboratories: yet, this financial restriction was the only serious restriction determined by my academic affiliation. In this relative sense, I maintain HPS scholars to possess a relative freedom with respect to scientists.

[95] Chang claims that social studies of science do not operate as complementary science. This is because the sociology of scientific knowledge 'deflates the special authority of science as a whole by reducing the justification of scientific beliefs to social causes. In contrast, the aim of scepticism and anti-dogmatism in complementary science is the further enhancement of particular aspects of scientific knowledge' (Chang 2004, 248). As illustrated at the beginning of this chapter, I agree with Chang's diagnosis about the disinterest (at best) and unwarranted scepticism (in the worst cases) in scientific questions characterising the majority of STS research, which is exemplified the Latourian stance that 'we want to be at once more scientific than the sciences - since we try to escape from their struggles - and much less scientific - since we do not want to fight with their weapons' (Latour 1988b, 165; note the pseudo-evangelical 'we', meant to involve all scholars regarding themselves as contributors to STS, and the rhetorical use of war imagery). However, I do not take this tendency to be indicative of what social studies can and sometimes do contribute to science itself. As exemplified by Arturo Escobar's work on anti-essentialism in ecology (1999), Stephen Helmreich's anthropology of the Artificial Life community (1998) and Steve Fuller's controversial (and, unfortunately, largely unchallenged) defence of intelligent design (2006), social and ethical considerations can have great bearing on the scientific understanding of nature.

fairly specialised study of scientific understanding, are especially interesting to scientists themselves and thus should explicitly aim at contributing to their work.

I hope to provide an illustration of this latter claim in the conclusion of this dissertation, where I discuss the possible impact of my analysis on the ongoing scientific debate about how to use databases and simulations in model organism research. In attempting to confront this scientific issue, my work indeed conforms to the goals of complementary science by producing results that challenge – and thus, potentially improve upon – current scientific practice. The fashionable area of genomics constitutes another setting in which complementary science might operate alongside specialised, normal science. The main preoccupation of scientists working in this field is to acquire knowledge about the relation between gene structures and their functions in the cell. This is seen as an important first step towards uncovering the mechanisms governing organismal growth and development (and, eventually, controlling them). It is not a geneticist's job, however, to conduct research on the way in which the history of genomics shaped current understandings of what a gene is, on the alternative conceptual frameworks that could be used to think about genes, or on ethical concerns about the use of genomic knowledge in society (such as issues of privacy, patenting, governmental regulations over biotechnology and so forth). This is problematic, since there is much that a geneticist could learn from these complementary types of research. Some biologists indeed profited from philosophical critiques of the foundations of genomics in building the field of evo-devo, now one of the most exciting area of biology that counters many of the gene-centric tenets of classical genomic research. Further, it is often argued that scientists have at least some responsibility towards the social significance of their research. Much sociological research is relevant to their work in this sense, for instance by documenting how scientific authority is misused and its results distorted in the public arena. Better information about these mechanisms might encourage individual scientists (and scientific institutions) to take interest in the circulation of their results beyond the laboratory, participate in public debates and refrain from exploiting such misuse for their own private purposes (as is sometimes the case in corporate-sponsored biotechnological research, where principal investigators are free to protect personal financial profit over the validity and trustworthiness of their results).

Does the idea of philosophy, history and sociology as forms of complementary science then imply that practicing scientists should take interest in it at all times? I would argue that this is not the case. This is not because contamination between science and philosophy is impossible or useless (as repeatedly claimed by some of the contributors to the infamous Science Wars[96]). After all, a clear distinction between the two types of practices cemented only towards the end of the 19th century. Moreover, the scientific fruitfulness of mingling research with philosophical analysis is demonstrated by several crucial episodes in the recent history of science. Michael Friedman has cogently argued that intellectual exchanges between philosophers Helmholtz, Mach and Poincaré and

[96] Scientific hostility towards the co-operation of scientists and philosophers is tellingly exemplified by Steven Weinberg's 'Dreams of a Final Theory' (1994), where a whole chapter (titled 'Against Philosophy') is devoted to declaring philosophy useless for scientific purposes. For a rich examination of Science Wars discourse, see Koertge (1998).

mathematicians Riemann and Klein provided an ideal terrain for Einstein to establish the foundations for special and general relativity (a process that Friedman calls ‘communicative rationality’ in his 2001 ‘Dynamics of Reason’). Cooperation between philosophy and science features heavily in the history of biology, too. Apart from the well-known philosophical threads in Darwin’s ‘Origin of Species’, a telling example here is the Modern Synthesis achieved during the 1930s: most biologists now regarded as responsible for that fusion of Mendelian and Darwinian insights were also active contributors to philosophical debates about the implications of their scientific views (as in the well-known cases of Ernst Mayr and Theodosius Dobzhansky).[97]

From these examples it might be tempting to conclude, as Friedman does in his own fashion, that it is preferable for scientists to take interest in complementary science alongside their ‘normal’ research. While this might certainly be the case at specific junctures in a scientist’ career, I am not convinced by this claim. A constant and/or conscious mingling of scientific and philosophical analysis is neither always necessary nor sufficient to the advancement of science. The general goals of normal and complementary science might be the same. Nevertheless, the ways in which philosophers conceive of scientific matters might differ so widely from the scientific perspective as to be incomprehensible to working scientists (and vice versa). It might be that such a gulf of interests is filled in the course of time and that philosophers find ways to make their thoughts relevant to scientific practice. It might also be that the line of thought pursued by philosophers turns out to be sterile or entirely misguided, just as in the case of a scientific research programme not fulfilling expectations and preliminary aims. In the latter case, complementary science will not have fulfilled its goals in the positive sense of contributing to the advancement of science.

In fact, this positive fulfilment is not necessary, as long as complementary science has another, more negative but not less crucial, connotation: this is, in Chang’s words, its ‘reluctance to place restrictions on the range of valid questions’ (2005, 239). Surely not all questions are of interest to the practicing scientist, nor should they be. What complementary science hopes to offer is an intellectual space where alternative questions can be asked without immediately worrying about their short-term utility or social repercussions. Such worries are intrinsic to specialised scientific practice. This is, at least in part, a good thing: given the limited availability of resources for scientific research, research programmes should indeed be carried out under time pressure, especially when their results may prove fruitful to society or, in case of failure, might steal precious resources from other programmes (as is often the case in the biomedical sciences). There are darker sides to the situation, especially since specialised science is becoming increasingly product-oriented and sponsored by agencies that value the applicability of results over their epistemic value. Complementary science is essential in order to supervise and critique the effects of these circumstances: it is perhaps unreasonable - and certainly unrealistic - to require specialised scientists themselves to conduct such

[97] See Mayr and Provine (1982).

complementary research while burdened by tight deadlines and heavy responsibilities towards funding bodies, peers and the public.[98]

4.3 Complementary Science in Action: Interacting with the Biologists

After outlining the goals and values characterising HPSSB as a form of complementary science, it is time to turn to the practical, methodological implications of this position. In this section, I take my own research as representative for this framework and I expose some of the issues that emerged while carrying it out. In particular, I address the ambiguity brought by this approach to the relationship between myself, as HPSSB *researcher*, and Arabidopsis scientists and their practices as the *objects* of my research.

4.3.1 From Participant Observation to Collaboration

My research centres around two specific subgroups of the Arabidopsis community: The Arabidopsis Information Resource [TAIR], that is the laboratory that collects, organises and makes available online all existing data on Arabidopsis biology; and the Nottingham Arabidopsis Stock Centre [NASC], where Arabidopsis specimens are grown, labelled and distributed to researchers for experimental use.[99] I started my investigation of these sites by consulting two traditional sources of material for trained historians. The first was published literature, including scientific publications about the history of Arabidopsis research, the results achieved by the two research teams and their role in the broader Arabidopsis community, as well as online presentations of the main laboratories involved in Arabidopsis research, such as the NASC and TAIR home pages and the websites of leading figures in the field (e.g. the Meyerowitz Lab). My second source was the ensemble of community archives, conveniently stored within TAIR and NASC premises, documenting the processes through which these two research teams and their practices came to be. I quickly found out that published literature contains useful information about the ethos and directions publicly endorsed by researchers from the two sites, but largely obliterates the actual activities and tools characterising their research. The same is true of websites, whose function is more to advertise than to document the tasks carried out in a specific lab. For the purposes of making sense of how Arabidopsis scientists acquire understanding, this strategy alone would not do. Archival work presented me with an even more pressing problem: it would not yield meaningful information without the

---

[98] These remarks are not meant to depict normal scientists as entirely unreflexive, nor to sanction their lack of reflexivity when it occurs. There are notable exceptions to this trend (such as, for instance, John Maynard Smith, who has reflected at length on the relation between Darwinism and capitalism; 1979, 32) and it might even be that the 'exceptions' constitute the majority of scientists: unfortunately, we lack empirical data documenting the extent to which scientists' concern with the implications of their work influences their research activities. My arguments here are not directed at the personal stands of individual scientists, but rather to contemporary scientific journals and institutions, the vast majority of which is certainly not encouraging reflexivity and social concerns in science. I thank Henk van den Belt for encouraging me to reflect on this point.

[99] For detailed information about the history and function of these two loci of Arabidopsis research, see section 3.3.

guidance of the biologists who gave me access to the records. This is largely due to the digital nature of the archives in question. Documents pertaining to communities built in the computer age risk to be far too many and far too unorganised for an historian to examine them all. Email exchanges among founders of TAIR alone run in the tens of thousands, all preserved in chronological order in the same database. Since in this case software does not allow for word-searches or quick scanning of text, there is practically no way to distinguish the majority of entirely trivial or purely technical messages from the small minority of text that might be relevant to my purposes. In order to select material without wasting years on a single archive, it became necessary to rely on scientists both for locating documents and correspondence and for complementing it with additional oral information.

In the summer of 2004, I thus started to pursue what anthropologists would call ‘fieldwork’ experiences – that is, periods spent in visiting the protagonists of my study on their home ground and participating, insofar as possible, in their daily activities. Those comprised a visit to the Carnegie Institute of Plant Biology in Stanford in August 2004 (home to TAIR as well as to a large group of brilliant Arabidopsis researchers); participation to a number of biology conferences (concerning the set-up of gene ontologies, as documented in Chapter 5, and the use of models in biology); a visit to the NASC in May 2005 (where I learnt the rudiments of planting and harvesting Arabidopsis seed, which I discuss in Chapter 6); and several conversations with leading Arabidopsis researchers based in England and in the Netherlands. These experiences brought a wealth of material and insight without which this dissertation would hardly have been possible. Indeed, personal interaction with biologists turned out to be at least as relevant to my purposes as archival research and literature searches. Moreover, visiting the physical space in which Arabidopsis biologists work, watching them interact with each other and with their laboratory environment, experiencing the space and conditions under which they work as well as studying the materials that they used, have all proved extremely useful clues in deciphering the more or less tacit skills that they employ, their strategies to communicate and persuade each other and their concerns. ‘Being there’, as ethnographers would put it, is very important to an analysis of contemporary scientific practices. In my case, it was essential in order to make sense of the research practices of TAIR and NASC scientists, assess the extent to which these practices are influenced by the context in which they are carried out and, most importantly, verify how tight the connections are between the scientists’ ways of reasoning and their ways of acting.

As it might be expected in the light of my discussion of complementary science, I found myself to be closely allied with the scientific goals pursued by the scientists I was studying. Like them, I am interested in determining the tools and strategies that are most appropriate to facilitate a cross-disciplinary, integrated understanding of plant biology. I agree with many of the scientists that I interviewed that disciplinary specialisation in the life sciences is both necessary and useful; I also share their unease in admitting that there are no resources nor venues available where such results can be brought together to bear on issues of common interest. I certainly wish my work to be useful to scientists trying to achieve such common conceptual platforms. What is most interesting about this sharing of commitments is the extent to which it facilitated my inquiry. It was helpful at the practical level, as the common engagement with Arabidopsis research facilitated my access to both resources and willing interlocutors. Most importantly, it was helpful to experience for myself the extent to which adherence to scientists' goals, exposure to their social context and acquisition of some of their material skills increased my own understanding both of their work and of the phenomena that they study. These three elements are the ones that I single out and discuss in my conclusions as crucial conditions enabling a scientific understanding of natural phenomena. In an interesting reflexive loop, I thus extracted my epistemological analysis not merely through a rational process of reconstruction, but by understanding my targeted phenomena (Arabidopsis practices) in much the same way as Arabidopsis scientists understand plant biology.

Here is a first aspect of the ambiguity that HPSSB research, in its form of complementary science, brought to the relationship between me as a researcher and the communities on which I work: what constitutes an object of my study also constitutes my closest collaborator, as it helps me to acquire the experiences through which I can gain an understanding of it. Another way to read this phenomenon is to note that, apart from sharing interests and goals, the biologists and I have similar stakes in the research being carried out. Neither of us can claim to be impartial towards the results and significance of Arabidopsis research: I care about the trustworthiness and outreach of knowledge about plant development in much their same way. I invested a great deal of time and resources in studying the Arabidopsis community, which involves both a personal attachment to their endeavours and personal interest in their successes and failures, which might even impact my career and future employment (even if, admittedly, not to the same extent as in the biologists' case). In addition, I see reciprocal, embodied understanding of phenomena as one of the most mysterious and successful aspects of science, as well as one fraught with social and political value that should, in my mind, be made explicit and subject to public debate. This care and personal investment constitutes a strong basis for efficient communication.[100] It also motivates my shift from participant observation to outright collaboration with the scientists that I investigated.

4.3.2 A Step Back to Situate Knowledge

The emphasis on the value of collaboration is not, however, the only guiding principle of my research. As already mentioned in my introductory discussion of complementary science, there is a strong sense in which my approach to HPSSB requires a certain detachment from scientific goals, values and experiences. My practice is not the same as the practice of NASC and TAIR scientists, nor does it aim to be: it differs from normal science primarily to the extent in which I maintain a critical distance from their modus operandi, terminology, interpretations and commitments.

[100] I owe this recognition to the work on participative research carried out by Steven Helmreich in conditions comparable to mine. Helmreich's research on the Artificial Life community, as reported in his (1998), remains a model of how daily interactions and critical engagement with researchers might yield enriching suggestions on both the value of Artificial Life research and its future development.

These latter statements might appear to flatly contradict my previous considerations on communality of goals and experiences. Yet, this seeming self-contradiction represents the main strength as well as the main difficulty of this type of research. HPSSB contributes to scientific understanding by challenging science itself. A detailed HPSSB study of scientific methodology thus inevitably involves the attempt to criticise such methodology, its theoretical roots and the interests it incorporates from as many perspectives as possible. Maintaining a critical distance enables me to recognise and assess alternative points of view with respect to the ones expressed by the scientists that I interview. It allows me to assess the credibility of each claim, as well as of my own related experiences, with reference to a number of other perspectives. In short, it is necessary in order to situate the knowledge gathered via a specific actor and/or experience in its proper context, thus understanding its value and place in the broader landscape of research practices, multiple interests and social networks that makes Arabidopsis biology what it is.

How to reconcile this requirement with the advantages attached to collaborative participant observation? Further reflection shows how the two attitudes are not, after all, incompatible with each other. Taking a step back does not mean renouncing the commitments and beliefs that I share with the scientists that I study. I can better describe it as a willing 'suspension of belief'[101] that sustains the epistemological interpretation of the events experienced on fieldwork. There is a sense in which the HPSSB researcher is constantly of two minds: while extracting input from her objects of study by sharing their goals, context and experiences, she keeps looking for reasons to distrust that input and assess it relatively to all other relevant information in her possession, be it historical, social or scientific. This systematic debunking also applies to the researcher's own framework, which needs to be critically challenged and re-adjusted depending on her exposure to ever more varied alternative perspectives and interpretations of the phenomenon under scrutiny. This exacting combination of sympathy and mistrust stretches even further the above-mentioned ambiguity of the position of an HPSSB researcher with respect to her objects of study.

To clarify how this might work in practice, let me turn to my own attempts to situate experiences and information gathered during fieldwork. Consider in particular the process of selecting a restricted set of relevant actors on which to focus my inquiry. Such a selection is unavoidable. As mentioned in the previous chapter, Arabidopsis research is conducted in over 5000 laboratories spread across the globe. Depending on their location and training, Arabidopsis investigators differ in culture, expertise, skills, instruments and contexts – a variation that has direct consequences for the practices that they employ. The amount of different activities carried out in the Arabidopsis community is thus simply too large for an HPSSB researcher to study it comprehensively. I was therefore forced to select a set of actors and practices that would be especially suitable to my investigation of biological understanding.

[101] I paraphrase the expression 'suspension of disbelief', introduced by the poet S. T. Coleridge to describe the mechanism by which we come to trust fictional entities and images, in order to underline the contrast between the instinctive quality of that process and the rational nature of philosophical analysis.

One difficulty posed by this selection consisted in calibrating the amount of trust to be placed on my initial sources of information as well as my own interpretation of these sources. My shift of attitude towards the initiative named ‘The Arabidopsis Book’ [TAB] constitutes a good example of the need to maintain a critical perspective even towards my own intuitions. TAB is a collection of essays by influential Arabidopsis scientists that aims at summarising the knowledge hitherto obtained on the plant. It is published exclusively online, so as to be regularly updated with novel insights. Thanks to my philosophical training, which greatly emphasises (one might say, overestimates) the importance of written texts and explanations, I initially took this digital book to constitute a wonderful case of scientific collaboration towards the integration of data – and, therefore, a good case study for my research. I thus devoted much of my initial inquiries to TAB and planned to study interdisciplinary understanding in the community by interviewing the various authors of TAB pieces as well as the editor and some readers. This was only to discover, within scarcely four months of research, that few Arabidopsis scientists considered TAB to be of much interest and, even worse, that there was no actual collaboration underlying its publication. Experts in different fields would simply send in their review articles, while it was up to the reader to understand the connections among them (which cannot be done in most cases, as different articles contained entirely different techniques, perspectives and terminologies to illustrate the same phenomena). Recognising this required a great deal of detachment from my own initial perspective and from the information I had gathered from Arabidopsis researchers who were exceptionally enthusiastic (though, it turned out, not that well informed) about TAB.

My research focus then shifted to digital databases, of which TAIR is certainly the most interesting by virtue of its commitment to conceptual integration in plant biology, and to the establishment responsible for the distribution of Arabidopsis specimens, i.e. the NASC. As discussed in Chapter 3, both of these laboratories tuned out to be immensely more influential to the production and exchange of knowledge in the community, thus fitting my intellectual interest in activities that explicitly targeted the need to bring together knowledge concerning different aspects of Arabidopsis biology. In addition to these intellectual motivations, the choice of TAIR and NASC as my research objects was attractive for strategic and practical reasons. These two centres are arguably among the most powerful and well-funded sites for Arabidopsis research worldwide. They benefit from excellent material as well as human resources: in fact, association to these laboratories involved gaining proximity to some of the leading figures in the community as a whole (such as, in the case of TAIR, Shauna and Chris Somerville). As recounted in Chapter 3, these scientists effectively started research on Arabidopsis and hold great power, to this day, on the directions and means pursued by the community. Also, they are directly responsible for the creation, organisation and preservation of the archive material that I needed to consult. Working with them thus implied accessing a wealth of information about aspects of Arabidopsis research that I could not have gathered anywhere else. For all these reasons, choosing the powerful seems extremely advantageous to an HPSSB researcher, especially if tackling large-scale science as in the case of the main model organism communities. It is no wonder that the vast majority of the historical studies focusing on popular model organisms have selected powerful scientists as protagonists for their narratives.[102]

Despite these advantages, however, there are good reasons to refrain from using powerful actors in HPSSB research on 'big science'. Some unhelpful implications of this choice are already evident in my own analysis. Access to key players in the development of Arabidopsis research was gained at the expense of the so-called 'periphery': it did not leave me the time to investigate samples of the thousands of Arabidopsis laboratories that, rather than imposing research directions and resources on the rest of the community, keep developing experiments and increasing knowledge pertaining to their own specialised domain. This limited the scope of my investigations to laboratories that, like TAIR and NASC, focus on achieving an integrative biology of Arabidopsis – a goal that is philosophically stimulating, yet far less representative of model organism research than the specialised work done in the vast majority of Arabidopsis laboratories. This characteristic needs to be taken into account when evaluating the epistemological significance of TAIR practices with respect to Arabidopsis biology as a whole: and yet, without data to document how 'periphery-bound' biologists view and use TAIR results, a balanced evaluation becomes very difficult to achieve.[103] The recognition of the tension between 'centre' and 'periphery' in Arabidopsis research required me to assess my sources of information with an eye to their privileged location.

The challenge posed by this problem is enhanced by the very personality and talents often displayed by the scientists I worked with. Principal investigators working in the most prestigious scientific institutes worldwide are, unsurprisingly, strongly opinionated about which issues and events matters the most to an historical and philosophical reconstruction of Arabidopsis research practices. Charismatic figures such as Chris Somerville possess an uncanny and highly rehearsed ability to advertise their views and argue for their credibility – the same ability that helped them to rise to their current positions. While striving to diversify my sources (by interviewing several scientists at different locations, including Maarten Koornneef in the Netherlands and a number of British contributors) and to underline the role and interests of scientists that I mention, my emphasis on the thoughts and actions of powerful individuals certainly influenced my epistemological assessment. This influence is fuelled also by my awareness that disputing the perspectives of people whose knowledge and power so far exceeds my own carries its own risks and responsibilities (especially given that the same people are likely to act as peers to my work or to influence my professional credibility).[104]

---

[102] This is true of all accounts of model organism research practices, including Creager's on the tobacco mosaic virus (2002), Kohler's on Drosophila melanogaster (1994) and Ankeny's on C. elegans (1997). A possible exception is represented by Karen Rader's recent book on mice (2004), which, by virtue of its focus on standardisation rather than the animals themselves, embraces a larger diversity of perspectives.

[103] A detailed discussion of the epistemological significance of the tension between 'centre' and 'periphery' in Arabidopsis research can be found in section 7.1.

[104] This issue has been hotly debated especially within anthropology, since this field is growing interested in the study intellectual and political elites, rather than the poor, powerless and geographically remote peoples preferred by traditional anthropologists. This welcome change of interests (aptly dubbed 'studying up') is usefully discussed by Nader (1974) and Helmreich (1998, 25). Sociologist of science Michael Mulkay (1981) has also usefully critiqued the relation between sociologists of science and the scientists that

These elements point to the necessity of holding a pluralistic vision of Arabidopsis research, one that offers an overall view of the practices and beliefs employed within the community, but that, precisely for this purpose, differs from each of the partial perspectives offered by individual Arabidopsis contributors. Like the vision of my informants, my own vision is of course, situated, as it is grounded in a specific set of interests and principles. Yet, neither the recognition of situatedness nor my dependence on the information provided by my scientific collaborators constitute insurmountable obstacles to providing an enriching view of scientific practices – one that is broader, more systematic, more historically and sociologically informed and less tied to financial interests than the views produced by biologists themselves. It is in this sense that I propose to contribute to biology, while at the same time taking critical distance from its normal practices: collaboration and analysis can co-exist within the same approach, without necessarily undermining the epistemological value of my findings.

4.3.3 Inevitable Asymmetries and Valuable Ambiguities

HPSSB research offers descriptive and normative accounts of scientific practice that derive their usefulness from the *asymmetry* characterising the role of complementary scientists and the role of normal scientists. Complementary scientists are less knowledgeable than normal scientists about the actual content and techniques used in scientific practice – in most cases, they would not be able to perform them themselves, just like I would not be able to sequence an Arabidopsis gene or to expose a plant to chemicals so as to increase its mutation rate. At the same time, because of this specialised knowledge and interest, normal scientists are rarely receptive to the broader scientific, social and historical context of their research and to its significance towards the construction and interpretation of their results. The asymmetries in perspective, methods and background knowledge characterising the two types of science are as inevitable as they are helpful. In this context, it is a matter of course that a HPSSB researcher should maintain an ambiguous role with respect to the scientists that she is studying. This point is so well entrenched to HPSSB practice as to look trivial. Yet, it is important to recognise it explicitly, so as to discern its main methodological implication: that is, HPSSB should choose methods that feed upon, rather than repressing, this valuable ambiguity.

As a conclusion for this section, I shall exemplify this point by discussing the strategy I adopted in order to conduct and analyse my conversations with scientists. This was, namely, the choice of avoiding to tape my conversations with my interviewees, thus

---

they study as one of 'intellectual vassalage': analysts of science are over-dependent on their scientific informants for resources and material, thus making it difficult for them to provide critical evaluations of scientists' actions and beliefs. I see this concern as very much alive within HPSSB. Yet, I think that a way out of intellectual vassalage is precisely the possibility, by analysts of science, to distance themselves from scientific practices and situate their own interventions. Reflexivity, coupled with relative financial aloofness and skills honed to reflect precisely about this issue, might go a long way towards freeing students of science from the constraints imposed by their dependence on their objects of study, i.e., in most cases, powerful scientists.

renouncing the opportunity to transcribe these exchanges and storing them as written documents that could be accessed by any other interested researcher. This decision has been widely criticised by several sociologists who have examined my work. As they pointed out, it compromised the accuracy and trustworthiness of my finding, by leaving me with no 'objective' trace of my verbal exchanges with the scientists. While acknowledging the damage that this choice could inflict on research of types other than my own, I wish to defend it in relation to my brand of HPSSB research. What I have been trying to achieve through dialogue with biologists is not a collection of statements and quotes that could usefully support my philosophical argument. This would not have made sense in my perspective, given the gulf separating the use of terms and arguments in biology and in philosophy. Quoting a biologist in a philosophical text is likely to have misleading results: the actual words in the quote might match the interpretation and terminology provided by the philosopher, but there is no guarantee that that interpretation actually corresponds to the biologist's. Far from securing an objective sense of 'what scientists really think', quotes clipped away from the transcript of a conversation black-box the processual nature of the communication between interviewer and interviewee, that is the way in which they keep learning from each other and adapting to each other's mentality and terms. Acknowledging the asymmetry in training and methods that distinguishes complementary from normal scientists implies acknowledging that the two sides need to have extended conversations in order to recognise the communalities of their overall goals as well as the difference in their tools and patterns of reasoning. This difference in paradigms, so to speak, need not be incommensurable: in order to overcome it, starting a dialogue challenging the references and overall vision of the subject is much more useful than the analysis of an interview transcript.

Again, consider my actual experience. One of my most useful sources of knowledge and insight turned out to be the extended dialogues I had with the director of TAIR, Sue Rhee, during my four weeks of residence in her laboratory. Thanks to her interest in my project, I was given office space on the floor of the Carnegie Institute that is occupied by the TAIR team, thus managing to interview most participants to TAIR research in their own working environment. While single interviews with several collaborators clarified different aspects and perspectives underlying TAIR work, the frequent and in-depth conversations I had with Rhee shed light on some crucial differences between our approaches, uses of terms and points of reference when discussing what would appear, at first sight, to be similar themes. Our seven two-hour interview sessions, not to mention briefer exchanges in Carnegie corridors and coffee breaks, were instrumental to my understanding of the meaning that biologists, rather than philosophers or historians, attribute to statements that I had seen in various MASC reports and Arabidopsis publications. One instance is the notion of 'integration' itself. To me, that denoted the merging of knowledge gathered by different biological fields. To Rhee, integration indicates more specifically the fusion of data gathered at the genetic level with data gathered at the organism-level. Once uncovered, this difference made both my questions and her answers more intelligible (otherwise, I would have kept talking about what I thought as the merging of theoretical explanations, while she would have answered with reference to merging databases). It also boosted my insight in the relevance of database modeling to Arabidopsis research. Another example concerns the term 'model system'.

Rhee uses it in a very different way from its philosophical equivalent[105], namely, as referring to the model of a component of a model organism. Further, she makes a distinction between this term and the term ‘model organism’, which denotes an organism as a whole to her. Recognising and understanding this difference forced me to revise much of my epistemological assessment of model organism research, as will become clear in Chapter 6 where I discuss the status of Arabidopsis as a material model.

Taping my conversations with Rhee and other scientists proved a practical obstacle to such meaningful communication. This was as much the result of the location and discontinuity characterising my exchanges[106] as it was a consequence of scientists’ uneasiness with being recorded and eventually misquoted. In the case of group meetings among scientists, this concern easily gives way to involvement in the discussion. In the context of a one-to-one interview, by contrast, the interviewee is rarely comfortable with the thought of being recorded. I noted such a stark difference between interviews carried out with or without the recorder, that I stopped recording altogether and started to take notes instead. Hence, my adoption of a methodological strategy that sacrificed the accuracy of my records on the altar of meaningful communication. Rather than looking for specific quotes, I have been trying to exchange ideas and get the gist of the researchers’ outlook, of their own interpretation of what they are doing. This helped at the practical level, too: taking notes during interview gave people time to think and the impression of being taken seriously, while maintaining a useful tension throughout the exercise (taking out my notebook would be a signal that ‘things are getting serious’, thus favouring a smooth shift from amicable, preliminary chit-chat to a focused discussion).

This process of communication wonderfully exemplifies the advantages of the characteristic ambiguity of HPSSB research. Meaningful communication can be achieved by agreeing on the meaning of terms and references used, by sharing experiences that allow for a reciprocal understanding of one another’s research and by meeting the interviewees on the common ground of a commitment towards increasing current scientific understanding of plant biology. At the same time, dialogue is fuelled by my ability, as a researcher gathering evidence on both scientific and social aspects of Arabidopsis research, to assess the situated nature of the views of my interlocutor.

4.4 Beyond the Particular: A Matter of Concerns

My combination of philosophy with history and sociology of science yielded a hybrid body of work. While certainly unable to tackle each relevant issue to the level of specificity required in each discipline, I hope that it goes some way towards breaking free of their insular limitations and thus address the broad issue of understanding with the

[105] Philosopher-historian Hans Rheinberger has introduced the very broad definition of experimental system as ‘the smallest functioning unit in science’ (1997, 306).
[106] In several cases, the time and circumstances of interviews were heavily constrained by the extremely busy schedule of the scientists I was interviewing. Interviews had to be conducted whenever convenient to the subjects. This meant, when I was lucky, over meals or coffee; otherwise, I would be discussing the meaning of experiments, the epistemic status of models and the ethos of the Arabidopsis community while hitchhiking a ride home or helping to transport material around the campus.

flexibility and open-mindedness that such an enterprise requires. The methodological choices and experiences accounted for in the previous section have hopefully left readers with a sense of how I accumulated evidence for the present piece of research, as well as with some motivations for these choices. I tried to emphasise how the goals and values underlying my research have informed my methodological decisions. However, rather than promoting a specific set of procedures as most appropriate way to pursue my goals, I have outlined the ambiguities intrinsic to HPSSB research in its complementary mode. In my view, there is in fact no precise course of action that might secure credibility and fruitfulness to this type of research. That is not to say that an HPSSB researcher has no methodological standards whatsoever. These standards are best expressed as a set of key concerns, which the researcher has to respect as much as possible while carrying out her inquiry. The precise significance and practical methodological implications of abiding to these concerns need to be established with relevance to the specific context and goals of such inquiry.

I shall now briefly discuss the three concerns that I think best constrain the space for action and interpretation within which an HPSSB project should be positioned. The first concern is *empirical adequacy*. In her interpretation of the cases that she studies, the HPSSB researcher strives to be as faithful as possible to her empirical findings, thus avoiding to distort and/or invent evidence in order to defend her claims. There should be no stance or interpretation of scientific practices and findings that an HPSSB researcher is not prepared to modify or even reject in the case of it not fitting the reality of what is observed. Efforts to provide empirically adequate descriptions of scientific practice should extend to both its (scientific, institutional, social, economic and political) context and its results (that is, the scientific knowledge that is gathered).[107]

The second concern on my list, *generalisability*, is indeed required to act as an antidote against such confinement. A well-known interpretation of this notion is that of generality as an absolute attribute of philosophical views, which denotes their universal applicability. This qualification of generality is mostly associated to rationalistic philosophy - even some social sciences occasionally aspire to it. Here I want to focus on a different interpretation of generality, one that is relative to the context in which this notion is used. This is the idea of generalisability in the sense of multi-locality[108]. According to it, a scientific claim is the more general, the more local contexts it can be applied to. These local contexts need to be other than the ones with reference to which the claim was originally developed. In this sense, the notion of generalisability is almost synonymous with the one of applicability: yet, it differs from the latter term insofar as it

[107] This apparently trivial point needs to be mentioned, given the above-mentioned tendency within some branches of social studies to exclude the intellectual goals and fruits of scientific research from their analysis of science. This first concern is central to historical and sociological methods of inquiry and needs to be integrated within (empirical) philosophy without, as I noted in my introductory section, necessarily confining the scope of philosophical analysis.

[108] Inspiration for this view is provided by Hans Radder's ideas about the non-locality of the results of scientific experiments (1996); and by Rachel Ankeny's arguments about the comparative nature of generalisations in the biomedical sciences, a body of knowledge which according to her is constructed on the detailed description of specific cases, which are then used as 'index-cases' for comparison with similar occurrences (2005).

includes both potential and actual applicability (in other words, one can reasonably suppose that a general claim might be applicable to a given context, unless such applicability has been empirically disproved).

The third concern of interest to HPSSB research is the one of *strategic reflexivity*. Reflexivity broadly refers to the awareness that a researcher is supposed to cultivate about the impact of her own beliefs, actions and reasoning on the data that she gathers and on her ways of interpreting them. As I argued when reflecting on the similarity of goals and concerns between my work and the biologists' work, active involvement in the issues to be researched might be the only way to produce historical, philosophical and sociological analyses. As participant observation gives way to collaboration, there is no denying the significant role played by the researcher in shaping the very reality that she studies. A complementary scientist actually aims at acquiring influence – thus offering, she hopes, a constructive contribution – on her objects of study. The reflexivity characterising an HPSSB researcher is therefore of a strategic type, since being aware of her impact on scientific practice is not just a means towards evaluating the credibility of her narrative, but also a means towards assessing her own success in pursuing the goals of complementary science.

My emphasis on concerns, rather than actual courses of action, highlights the extent to which HPSSB research methods are tied to the contingencies of a specific situation and the preferences of individual researchers. In this purely procedural sense, HPSSB does indeed produce a philosophy of the particular. On the other hand, my account of the goals of HPSSB research about contemporary science highlights two of its crucial components: (1) participative experiences alongside scientists, which aim at understanding their practices, allegiances and modes of communication and which are guided by the above-mentioned concerns; and (2) reflective moments of rationalisation in which those experiences are assessed and conceptualised relatively to background knowledge and beliefs. In the case of my own research, these two components produced a rational reconstruction of my participative experiences. This reconstruction was guided by my intellectual interests and methodological concerns, as well as by the abstract concepts and categories through which I chose to structure my analysis (such as the notions of theory, model, commitment and skill).

The resulting narrative yields a series of claims about the characteristics of biological understanding in the case of Arabidopsis research. These claims are not *assumed* to be transferable to contexts other than the Arabidopsis community: HPSSB research in its complementary mode does not, a priori, pretend its results to be universally valid. Nevertheless, they certainly have the *potential* of being applicable to other contexts: they are general in the sense of having the potential of being multi-local. Thinking of generality in terms of multi-locality means making it subject to empirical research: we can test the generality of a claim by verifying how it applies, if at all, to contexts different from the one that it has been developed to describe (that is, by evaluating its empirical adequacy in different sites). This view does not imply that those claims should maintain a fixed interpretation independently of the context to which they are applied – which is in itself advisable, since the meaning of concepts, as well as the nature of the relations among them, may well vary when travelling through different environments. This version of HPSSB research does imply, however, that uttering general claims is perfectly compatible with the study of specific cases. This is an important result, since the formulation of general claims provides a standpoint for further work and for the development of alternative (or even incompatible) interpretive frameworks to the one I proposed. In the context of HPSSB research, coining a claim means accounting for its local origins, motivations and development. As I hope to have shown, such information in turn facilitates, rather than hampers, the transferability of that claim to other contexts.

Chapter 5. Using TAIR to Understand Arabidopsis: Theoretical and Performative Skills

Ultimately, our goal is to provide the common vocabulary, visualisation tools, and information retrieval mechanisms that permit integration of all knowledge about Arabidopsis into a seamless whole that can be queried from any perspective

Sue Rhee Website, accessed November 2003

As recounted in section 3.3, The Arabidopsis Information Resource [TAIR] was set up by the Arabidopsis Steering Committee at the end of the 1990s in order to gather together and organise the various types of data progressively accumulated on Arabidopsis biology. The main practical purpose of this exercise is to facilitate Arabidopsis research, by making data about several aspects of Arabidopsis biology freely available to any biologist who might require access to them.[109] As I shall illustrate, the pursuit of this goal alone represents a major scientific feat, since it is very hard to classify and order the ever-growing mountain of data gathered by all contributors to the Arabidopsis community. Yet, what makes TAIR unique among the existing biological databases devoted to Arabidopsis is not the focus on storing and retrieving data. Rather, it is the realisation that compiling diverse data into one resource requires reference to an integrating framework. This realisation determined a crucial expansion of TAIR goals: from the practical necessity of finding ways to archive Arabidopsis data to the ambitious wish to construct tools 'that permit integration of all knowledge about Arabidopsis', in the words of TAIR Principal Investigator Sue Rhee (above).

In this chapter I intend to focus on the intertwinement of practical constraints and conceptual judgements underlying the making of TAIR (section 5.2) and on the nature of the integrating framework adopted by TAIR to arrange data and thus 'integrate knowledge' (section 5.3). Further attention will be bestowed on the embodied knowledge needed by scientists working at TAIR in order to build the resource (section 5.4). As it turns out, both the construction and the use of TAIR require a series of *epistemic skills* (section 5.5), without which it is impossible for biologists to use TAIR in order to further their understanding of Arabidopsis biology. This determines a tension between the goals of the TAIR project, since the emphasis on making the resource easily accessible does not square with the need to acquire additional skills in order to use it effectively. I start my analysis by illustrating what the TAIR project actually achieved up to now. In other words, it is high time to discuss what TAIR actually is and how it works.[110]

---

[109] See thee report compiled by the Multinational Arabidopsis Steering Committee in 1999.

[110] As specified in Chapter 4, the material that follows has been gathered through extensive interviews with members of TAIR, including Sue Rhee, Eva Huala, Leonore Reiser, Danny Yoo, Doug Becker, Katica Ilic and external advisers Chris and Shauna Somerville. The interviews were conducted during a month-long visit to TAIR in August 2004, during which I was granted a working space within the Carnegie Institute as well as access to TAIR archives.

5.1 What *Is* TAIR?

To Arabidopsis biologists, TAIR is primarily a website (www.arabidopsis.org), whose homepage is reproduced in figure 5.1. The website functions as a gateway to the various search and visualisation tools elaborated by the TAIR team to make Arabidopsis data retrievable by TAIR users. It also provides abundant information about how databases are made, how they should be consulted and which types of data are included in the main TAIR data set. Last but not least, the website allows users to order stocks of Arabidopsis seeds from the main Arabidopsis Stock Centres.

Figure 5.1 - TAIR Home Page (accessed 10 March 2006).

The TAIR website provides access to several databases, each of which collects and classifies a specific set of data. Examples of these databases are MapViewer, which allows access to various types of mappings of Arabidopsis chromosomes; AraCyc, which contains data about biochemical pathways characterising Arabidopsis cellular processes;

and BLAST, which displays Arabidopsis metabolic cycles.[111] All data presented in this way are stored in a central archive hosted by the TAIR server in Stanford.

By clicking on one of the hyperlinks for databases (that is, one of the items classified in the homepage under the heading 'analysis tools'), users are asked to select and specify parameters for their query, thus formulating their search according to the framework adopted by the database that they have chosen. For instance, they can choose to ask for data related to a specific gene marker on Arabidopsis chromosome 1. To do that, they would click on MapViewer, choose chromosome 1 among the five Arabidopsis chromosomes and fill in the name of the locus of interest (e.g., genetic marker SM39_256). Once the query is compiled, another click of the mouse is sufficient for the software to locate the relevant data and display it to the user.

Notably, TAIR offers a variety of parameters in which the same query can be pursued, so as to maximise the efficiency of any given search. Also, it proposes several ways in which the results of a query can be displayed. So in the case of the above query about genetic marker SM39_256, the user can view results in five different formats, including a classical genetic map, an AGI map (that is, the type of sequencing adopted by contributors to the AGI project) and a RI map ('recombinant inbred', reporting information about genetic markers derived by crossing two different Arabidopsis ecotypes).[112] These visualisations of Arabidopsis data are crucial to the functioning of TAIR. They allow the researchers working at TAIR (which I shall refer to as 'TAIR curators') to organise the sea of data stored in the central server, so as to make them easily retrievable. Most importantly, these images do not simply facilitate the users' access to the data: they incorporate a specific interpretation of the data, which is used to select data in the first place, to connect them to a specific biological context and to signal their relevance to specific queries and types of research. In this chapter, I wish to focus on the interpretive component of TAIR and on its implications, especially concerning the prospective usefulness of TAIR to the Arabidopsis community. Through a step-by-step reconstruction of the main phases in the construction of TAIR, I will show that TAIR images are constructed around a theoretical interpretation of the data that they display: TAIR curators use a *network of concepts* known as Gene Ontology [GO] in order to classify, organise and display available data on Arabidopsis.

The construction of TAIR databases turns out to be an extremely complex process, which involves making significant assumptions about how to model data collected on Arabidopsis. Representing available data on the plant implies some degree of interpretation of those data and their significance towards Arabidopsis biology as a whole. Reference to a conceptual framework is unavoidable when trying to construct such a complex set of databases and images, as decisions about how to organise data are never entirely neutral and/or objective, but rather they are geared towards the fulfilment

[111] For scientific details on TAIR and its component AraCyc and MetaCyc, see Rhee (2000), Huala et al (2001), Reiser et al (2002), Garcia-Hernandez et al (2002), Rhee et al (2003), Mueller et al (2003), Krieger et al (2004).
[112] http://www.arabidopsis.org/servlets/TairObject?accession=GeneticMarker:2225346, accessed 20 March 2006.

of specific goals. This is especially true in the case of TAIR, whose long-term goals are to make Arabidopsis data not only accessible, but also comprehensible to biologists. As remarked in Chapter 3, TAIR was set up as a digital platform on which different data sets could be confronted and eventually integrated, so as to facilitate the understanding of Arabidopsis biology as the biology of a whole organism. It is in this sense that TAIR constitutes an extremely interesting, innovative bioinformatic tool. At the same time, the conceptual framework required for TAIR to enhance biologists' understanding of Arabidopsis biology is precisely what makes it problematic for them to use the resource efficiently.

I intend to show that users can access the visualisations of data provided by TAIR, and thus use them to facilitate their own research, *only if* they are acquainted with the theoretical framework through which they were produced. This presents TAIR with a serious problem. The conceptual framework elaborated by TAIR curators (that is, the GO network) is tightly embedded in a specific, gene-centred epistemic culture. Indeed, as I shall illustrate, understanding TAIR images requires possessing a specific set of interpretive skills, which corresponds to the one used by TAIR curators in creating the images. This finding contrasts with the function that TAIR was intended to serve, that is, making Arabidopsis data accessible to all interested biologists, regardless of their individual skills, convictions and expertise.

Before proceeding with my analysis, let me emphasise one important aspect of the scientific activities and results that I am about to discuss: its relatively recent birth and its unpredictable future. The TAIR team has started its research activities in 1999, thanks to a first research grant covering the period from 1999 to 2004. The continuation of the grant was conditional on the quality of the results obtained in that first trial period. An assessment exercise that took place at the end of 2004 found TAIR work to be extremely successful, as a consequence of which the project has been granted further funding until 2009. The short life of TAIR is a factor that needs to be kept in view throughout my analysis. This is, first of all, because TAIR work is still in an experimental phase. This makes it an exciting site at the cutting edge of contemporary biological research, as well as an excellent place for a philosopher to work in collaboration with biologists: TAIR curators welcome all insights and suggestions for improvements on their current work. Second, the short life of TAIR means that its results are still relatively limited, especially since, as I am about to document, a long stretch of time has been devoted to building the foundations (technical as well as conceptual) on which to construct this resource. Up to now, TAIR has constructed visual representations of data concerning only the lowest levels of organisation in Arabidopsis biology. This initial focus on genomics and molecular biology was determined by the material at hand, that is, by the abundance of gene-level data gathered in the 1980s and 1990s through generously funded projects such as the Arabidopsis Genome Initiative. Given the urgent need for databases that would store and organise all that information, the first TAIR grant contained short-term goals that were explicitly targeting gene-level information and allowed for the inclusion of higher-level data (for instance, data pertaining to cellular organisation and ecological adaptability in Arabidopsis) only on the basis of gene-level information (Garcia-Hernandez et al 2002).

This might appear a meagre scope, when confronted with the ambitious integration of data ranging from the ecological to the genomic level advertised by TAIR as its long-term goal. In fact, I shall later criticise TAIR as ending up, despite its proposed goals, enforcing a gene-centric view of Arabidopsis biology. What I want to emphasise in this section, however, is the extent to which the existing databases and relative visualisation tools hitherto elaborated by TAIR curators constitute a major achievement in the context of Arabidopsis research, as well as in the broader landscape of model organism research. The work of TAIR curators involved extensive trial-and-error phases in collaboration with similar research teams in other model organism communities, consultations with a variety of disciplinary experts and bioinformaticians, as well as crucial decisions on how to go about interpreting and representing Arabidopsis data. Further, there is a sense in which TAIR personnel is aware of the shortcomings intrinsic to their approach: Principal Investigator Sue Rhee and Managing Director Eva Huala insist that the current choice to organise the database on the basis of genomic data is purely pragmatic, since the ultimate TAIR aim remains to capture information pertaining to all levels of biological organisation, from the genomic to evolutionary/ecological levels.

In the years to come, TAIR curators will have the chance to prove whether their approach and methods can be stretched as far as to encompass such different data sets. It is impossible to predict whether their research will fulfil its original goal and how biologists within and without the Arabidopsis community will profit from it: I shall therefore abstain from speculations in this respect. This uncertainty notwithstanding, there is a strong sense in which no scientist (or philosopher) is more aware than TAIR curators of the difficulties involved in creating a digital resource encompassing a large quantity of data of widely differing origin and type. The next section is therefore dedicated to reconstructing the work conducted at TAIR and, as discussed in Chapter 4, is based on my interviews with TAIR curators, relevant scientific publications and my observations on site.

5.2 The Making of TAIR

5.2.1 Design: Vision and Realisation

The first (and, in the view of TAIR researchers, most important) phase in the construction of TAIR databases is the one of *design*. By design, TAIR curators refer to the preparatory work needed to publish databases on the TAIR site. This includes learning to use appropriate tools, clarifying ideas about which types of databases are needed and selecting what needs to go into the databases, and in which form. For the purposes of my analysis, I divide this cluster of design-related issues into two sets of research activities, which I call *vision* and *realisation*. Interestingly, these two aspects of design define a division of labour among TAIR personnel: TAIR curators, who are all trained biologists with an interest in the scientific value and validity of TAIR, concern themselves mostly with the aspect of vision; while TAIR programmers, that is the IT engineers hired by TAIR to provide technical support to the curators, are in charge of the aspect of realisation.[113] I now discuss each of these aspects in detail and look at how they relate in practice.

Visions for the User

As remarked above, TAIR databases need to be easily readable and accessible to the user. For TAIR curators, this involves thinking about (i) how data can be visualised without loss of information and (ii) whether the proposed options for data visualisation would enable prospective users to easily locate and access the information they want. This latter aspect is especially important, as TAIR sponsors and scientific patrons conceive of the TAIR project as a service to the Arabidopsis community as a whole. TAIR director Sue Rhee agrees that the primary goal of TAIR consist of what she calls 'facilitation', that is, of making the acquisition of information about Arabidopsis as effortless as possible for all potential users (that is, biologists likely to need information to further their own research).[114] The search tools contained in TAIR are therefore expected to please Arabidopsis biologists, thus encouraging them to benefit from the resource by making regular use of it. TAIR curators have taken this expectation to imply that the first step towards the construction of TAIR should be discussing what a database should look like, in order to satisfy as many of its potential users as possible. Agreement on a specific template for a database, they reasoned, would allow them to set out the process of actually producing the database. These templates are what TAIR curators refer to as their *visions* of what the user wants from a database. I thus borrow the term and its definition directly from them.

The three TAIR curators responsible for creating visions are Sue Rhee, Eva Huala and Lead Curator Margarita García-Hernández, with occasional input provided by members of the TAIR executive committee (including Chris Somerville).[115] Rhee firmly believes in the importance of spending far more time on the planning and preparation of visions than on their actual realisation, as she emphasises that 'it takes more time and pain to 'redo' than start slow'.[116] Rhee and Huala point to two main difficulties in their work towards creating a vision of the database.[117] The first difficulty was, unsurprisingly, information management. The early days of TAIR research involved frantic consultation of all types of literature dealing with this subject, in the hope of finding suggestions about how to visualise the most diverse information in the simplest possible way, without losing sight of its richness and of the diversity between features and sources for each datum.[118] TAIR curators confronted this problem by deciding to create several different

[113] TAIR members have published their views on the effectiveness of collaboration between software developers and biologists in Weems et al (2004).
[114] Interview with Sue Rhee, 12 August 2004.
[115] As far as the elaboration of the vision is concerned, Huala cannot recall any other external influence, neither from within the Carnegie nor from other members of the Arabidopsis community.
[116] Quote taken from a presentation given to NSF by Rhee in May 2004.
[117] Interview with Eva Huala, 13 August 2004.
[118] Huala points to Tufte (2001) as a main source for inspiration. Interestingly, she drew inspiration particularly from historical examples contained in the text, such as for instance graphics displaying information about Napoleon's campaign to Russia.

databases, each of which would provide a different perspective on Arabidopsis biology. They therefore devised visions for a database containing data about the location of genes on Arabidopsis chromosomes; another displaying data about gene expression; another focused on data about metabolic cycles; another about biochemical pathways; and so on. The possibility to gather data about the same phenomena from different perspectives, they reasoned, would indeed maximise the information available to users, without losses in the accuracy or the richness of data. Further, within each database users would be allowed to formulate their query in different ways: they would be able to choose among different parameters, as well as different ways to display the results of their search (for instance, when searching a specific gene locus on a chromosome, current TAIR users can view their results in the form of a genetic, physical or sequence map[119]).

A second difficulty consisted in trying to imagine what the user wants, since, as TAIR curators know all too well, 'the' ideal user of TAIR does not exist: TAIR aims to reach several types of scientific audiences, ranging from 'the' developmental biologist to 'the' technician, 'the' molecular biologist and 'the' theoretical biologist (not to mention the differences in epistemic culture to be found within and across those very categories). In order to confront this problem, TAIR curators drew insight from their own experience as experimental biologists specialised in different areas of Arabidopsis development.[120] As they told me, they tried to 'play the user': that is, they tried to imagine what an experimenter engaged in a number of research activities would expect from a database. This strategy proved to be extremely useful. For instance, Huala proposed, on the basis of her own experience as a developmental biologist[121], that the user should be able to 'fly into' the chromosome, i.e. to view and explore a three-dimensional representation of Arabidopsis chromosomes that is produced and constantly modified on the basis of incoming experimental data. This vision meant that TAIR should provide complex, three-dimensional visualisation tools that would allow users to 'click' on the image of a specific chromosome and see a representation of the inside of the chromosome. On the one hand, this representation would have to be realistic enough as to convey ideas about the actual structure and physiology of chromosomes; on the other hand, it would have to

---

[119] As recounted in the current TAIR site (URL: http://www.arabidopsis.org/mapViewer/help/tairmapa.jsp, accessed 14 March 2006), TAIR's maps are divided into three general types. *Genetic* maps are constituted by sets of 'markers' (i.e., specific areas of Arabidopsis chromosomes) that have been mapped by crossing parents with two different forms of the marker (in the case of many classical markers, these are the wild-type and the mutant) and scoring how many recombination events separate the marker from other markers on the same map. *Physical* maps are collections of overlapping clones that have been arranged into a tiling path based on either fingerprinting (digestion of clones with restriction enzymes and comparison of the fragment sizes) or hybridization. *Sequence* maps are sets of clones that have been sequenced or have been chosen for sequencing by the Arabidopsis Genome Initiative (AGI).

[120] While Rhee has a background as molecular biologist, Huala joined TAIR after years of training as a developmental biologist.

[121] The idea of 'flying into' (or around) objects of interest is certainly not new, nor is it recent. Such methods have been in use within engineering, for instance, since at least two decades. Huala herself was inspired by visioning a similar tool, elaborated by another research group to simulate a bacterial system. What I want to emphasise here is thus not so much the innovative nature of this idea, but the free way in which curators went about exploring possibilities, as well as their primary motivation while doing so: that is, satisfying the user.

contain specific references to the data from which the model was generated, so as to allow users to trace the sources and original context of the data.

Like other visions elaborated by TAIR curators, the idea of ‘flying into’ chromosomes became a strong heuristic in the construction of TAIR databases and eventually led to the creation of various types of chromosome maps (which, as in figure 5.2, represent chromosomes as a line on which genes and other markers are located). It also encouraged curators to think about spatial representations for information that is not normally visible in a real chromosome, but that is best visualised when using different colour patterns and three-dimensional effects. A relevant instance in this respect are the images provided by VxInsight, a tool borrowed by TAIR from Sandia National Laboratories. This database visualises gene expression patterns (a notably abstract type of data) as mountain terrain maps, where the red peak of each ‘mountain’ signals a cluster of genes characterised by a particularly pronounced expression pattern (figure 5.4).

Figure 5.2 - Map of the five Arabidopsis Chromosomes. The map follows the suggestion of users ‘flying into’ the chromosomes, by letting users zoom in specific areas of the chromosomes and seeing what gene clusters they contain (for instance, on this map a specific gene locus, named At1g01010, is signalled).

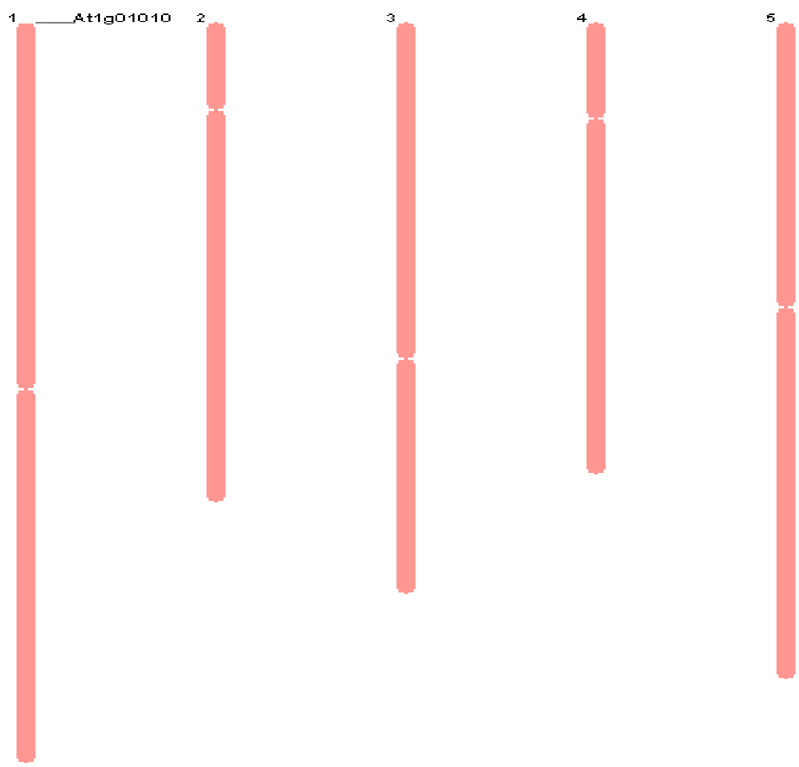


Figure 5.3 - Close-up view of a specific area of Chromosome 1, with indications about how to change display parameters and area of choice.

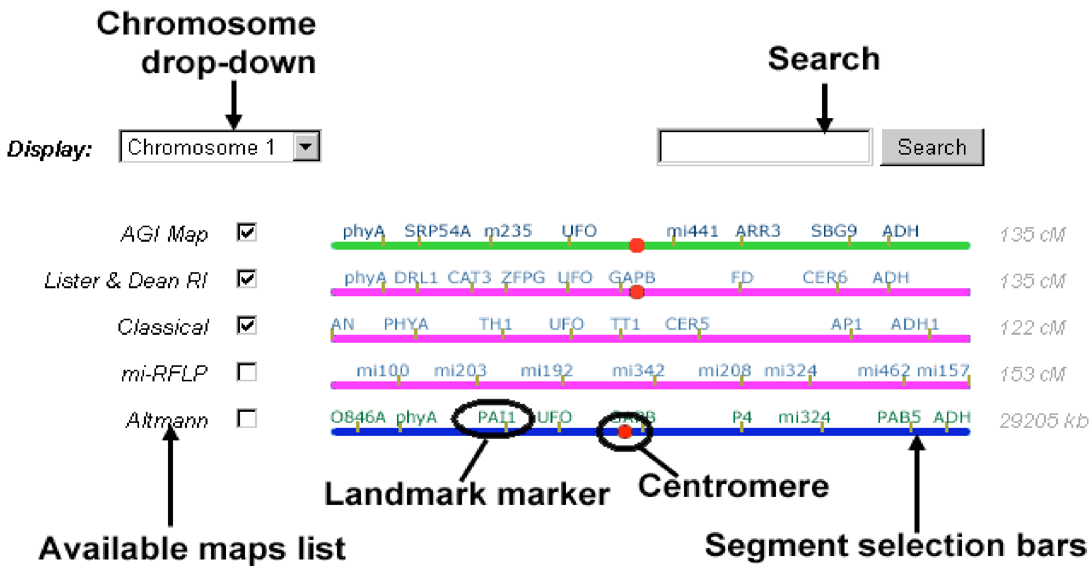
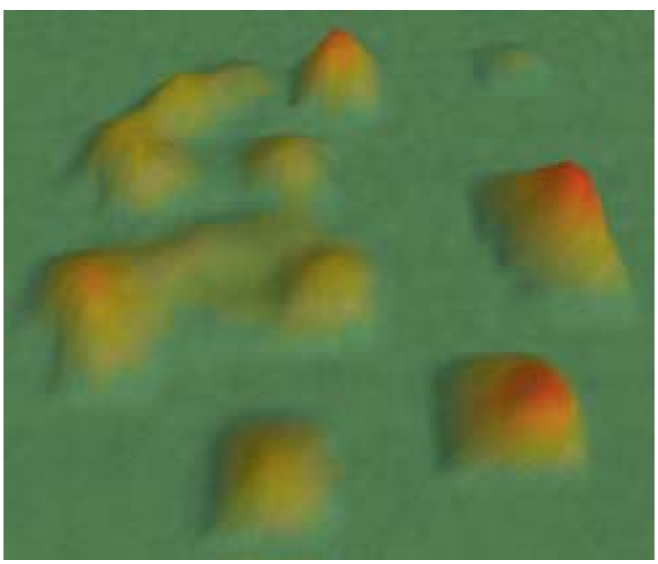


Figure 5.4 - VxInsight view of patterns of gene expression as physical mountain terrain.



Ideally, the best way to obtain feedback on the visions underlying the visualisation tools above would have been for TAIR curators to ask the opinion of experimental biologists with no direct involvement in TAIR (that is, the actual users for which those tools were being devised). Obtaining input from users proved, however, impossible, at least in the early stages of TAIR research. This is because providing input requires extended familiarity both with the new kind of database that TAIR was supposed to devise and with the goals and tools characterising the TAIR project. Biologists who were ignorant of the content and opportunities afforded by such a database were not in a position to give feedback on how best to construct it.[122] Rather than asking biologists for direct feedback on the vision of the database, therefore, curators started asking for queries that would be likely to be submitted to a database such as TAIR. The crucial question for the TAIR team then became: can we answer this query with the current vision? By asking this question, TAIR curators effectively brought together their concerns about information management and user-friendliness, thus elaborating visions for easily accessible, clear and yet rich databases.

Realisation: Java Software and the Design Loop

Once curators reach agreement on what they would like their databases to look like, the vision thus elaborated becomes a heuristic guide in making actual databases. In fact, thinking about design means also to conceive of ways in which such 'ideal' visualisations can be produced in practice. This is the aspect of *realisation*: appropriate tools and methods need to be found and/or built in order to realise the vision as accurately as possible.

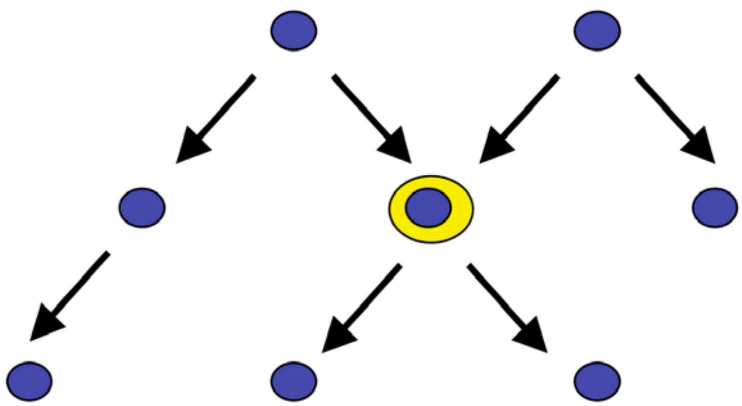
The realisation of the visions agreed upon by TAIR curators meant finding software that would support them, while at the same time being simple enough for biologists to learn to work with it. As I pointed out earlier, the task of proposing appropriate programmes and modifying them in order to fit the curators' demands fell on the hands of the IT engineers hired by TAIR for technical support (to whom I shall refer as 'TAIR programmers'). Interestingly, TAIR programmers have little if any training in biology to begin with, while TAIR curators have only a minimal training in programming. Rhee praises the advantages of this situation, as, in her words, it prevents either side from 'making assumptions' about the demands and goals of the other: motivations and reasoning underlying each choice have to be made explicit, so as to bridge the gap in expertise. At the same time, this situation also requires the two groups to collaborate closely over long periods and learn some of each other's skills in order to obtain acceptable results.

According to TAIR programmers, Java programming represents the best combination of effective rendition and ease of use for the purposes of TAIR. The basic set-up of Java programming requires the a priori selection of a set of objects and a set of relations among them, which are generally referred to as 'family relations'. These objects and relations are arranged into a so-called DAG structure (a 'directed acyclic graph'; see figure 5.5), where each object can be both the 'child' and the 'parent' of other objects. Note that the nature of these family relations is purely structural: it is a way to describe formal relations. Even the analogy with family ties, on which this terminology is based, derives from the asymmetrical nature of these relations: while any object can have

[122] As I highlight in Chapter 7, this might now be changing. Experimenters who are now acquainted with TAIR are expected to offer increasing feedback on the way in which the data that they produce are represented within the resource. This is because, as TAIR becomes a standard tool for publishing Arabidopsis data, experts will increasingly want their say about how data are stored and displayed. This feedback is welcomed by TAIR, since it relieves the curators from the great amount of work necessary to elaborate visions, while at the same time improving the quality and user-friendliness of the final product.

multiple parents and be parent of multiple objects, the same object can never both the parent and the child of another object (hence the arrows pointing only downwards in figure 5.5).[123] As we shall see, both of these features prove very important for the purposes of TAIR. Programmers refer to this approach as *object-oriented*, precisely because it involves choosing a set of objects around which all elements of the programme rotate.[124]

Figure 5.5 - Directed acyclic graph (DAG) used by object-oriented Java software. Each 'child' (i.e., object connected to other objects, here represented as a yellow/blue dot) may have multiple 'parents' (objects from which other objects derive).



To test the viability of this programming approach, TAIR curators present programmers with a series of mock-ups, that is, sketches of what the database should look like and how should it function, according to the visions previously elaborated. These sketches contain actual 'use cases', that is, visualisations of the specific elements that the users would expect to see in order to submit their queries and be satisfied by the results. Programmers then design appropriate software in order to match the user cases presented by the curators. This phase requires a high level of IT expertise, as it involves the re-programming of existing software (and sometimes even the creation of entirely new software). Once the software is deemed to be ready for use, programmers try to implement it on the user cases and pass it on to the curators so that they can verify how closely and accurately the implemented software matches their initial vision.

At this point, the aspects of vision and realisation (as well as the input of curators, on one side, and programmers, on the other) join into a *design loop*. If the implementation receives positive feedback from the curators, programmers make eventual minor

[123] The National Institute of Standards and Technology defines a DAG as a graph 'with no path that starts and ends at the same vertex' (http://www.nist.gov/dads/HTML/directAcycGraph.html, accessed 28 March 2006).
[124] According to Rhee, there are in fact not many viable alternatives to the Java object-oriented approach, which was immediately chosen by the TAIR community as the obvious candidate for database construction (email correspondence 19 June 2006).

modifications and improvements, after which the software is tried in a staging environment (that is, a set-up that mimics the insertion of the resulting database into the official TAIR website and thus allows to test the software in conditions that are closer to the real conditions of routine use). Approval at this stage means that the software, and corresponding database, are ready to be issued to the wider public. In most cases, however, the initial implementation tests give unsatisfactory results, either from a technical (malfunctioning) or from a biological perspective (misrepresentation of the curators' vision). Hence, the user case goes back to the programmers, who attempt a different IT strategy to obtain more appropriate results.

This loop of feedbacks between curators and programmers lasts as long as it takes for the realised database to match the vision proposed by curators. Of course, the technological constraints introduced by information technology imply that the vision itself has to be modified, sometimes radically, to fit a format that it is actually possible to produce and handle. Also, studying an actual prototype often allows curators to correct mistakes contained in the initial vision, which they could not recognise on the basis of pure speculation. Consider for instance the example of a design loop displayed below. On the basis of curators' vision about the information that the database should yield an answer to a query about a specific gene, programmers construct a database interface that looks like figure 5.6. Curators approve the technical accessibility of the prototype, which is indeed very easy to obtain by digiting the name of the gene (in this case, AT5G63210.1) into the appropriate search engine. However, looking at the actual prototype makes them realise that the initial vision lacked a crucial parameter: there is currently no information about the 'obsoletion status' of data (that is, the length of time since they have been inserted, as well as the relation of those data to more recent data). Thus, curators propose a corrected mock-up of the initial vision, where information about obsoletion is prominently displayed (figure 5.7).

Figure 5.6 - Prototype of database on gene AT5G63210.1 that lacks information about the obsoletion of data.

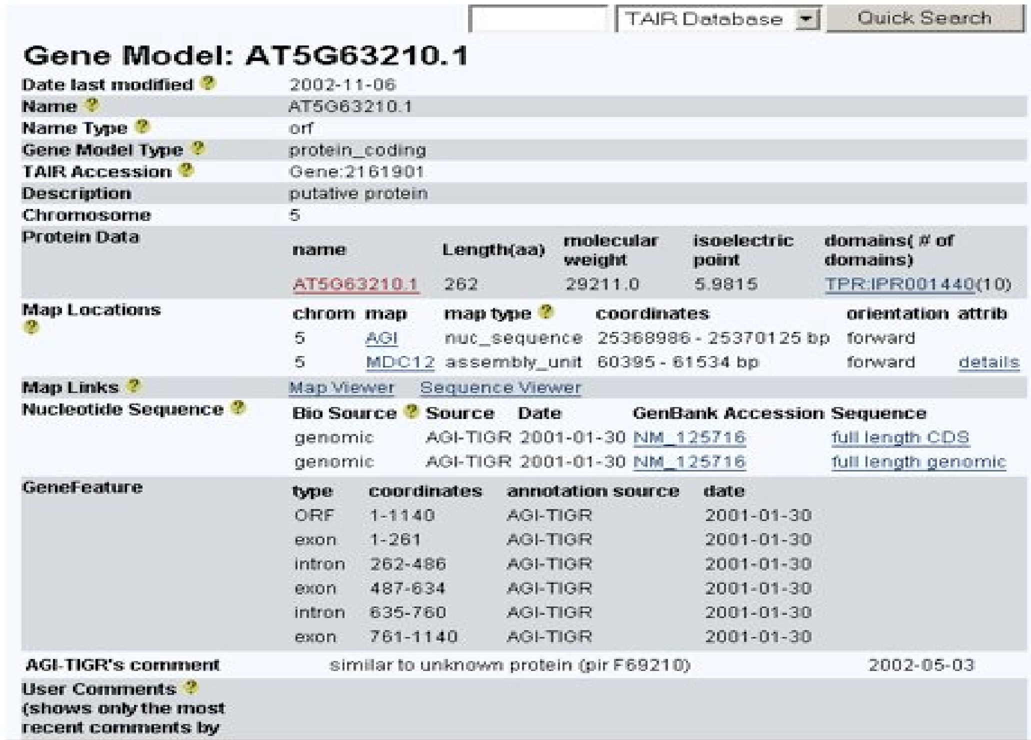
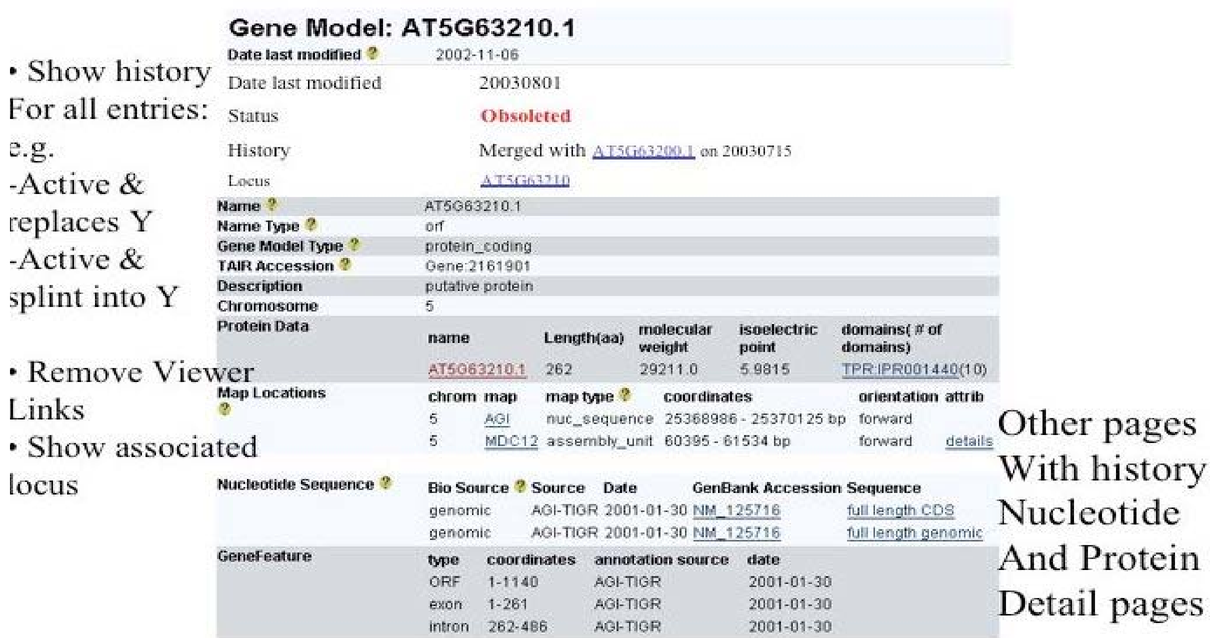


Figure 5.7 - The same mock-up after undergoing 'correction' by curators. This corrected version of the file, containing information about obsoletion and comments in the margin concerning data history and suggestions, is what programmers receive from curators during the design loop. It contains the material that they will have to work upon in order to realise the database according to the curators' vision.



5.2.2 Conceptual Framework: Gene Ontologies

My discussion of the interdependence of vision and realisation characterising the design of TAIR highlighted TAIR programmers' preference for the object-oriented approach supported by Java. It did not, however, explain why TAIR curators endorsed that choice as appropriate to the organisation and integration of biological data. To appreciate the curators' reasons for settling on this programming tool, we need to think about the actual phenomena that the data incorporated by TAIR are supposed to document: that is, the mechanisms, processes and components of plant development and molecular biology that are explored through the study of Arabidopsis.

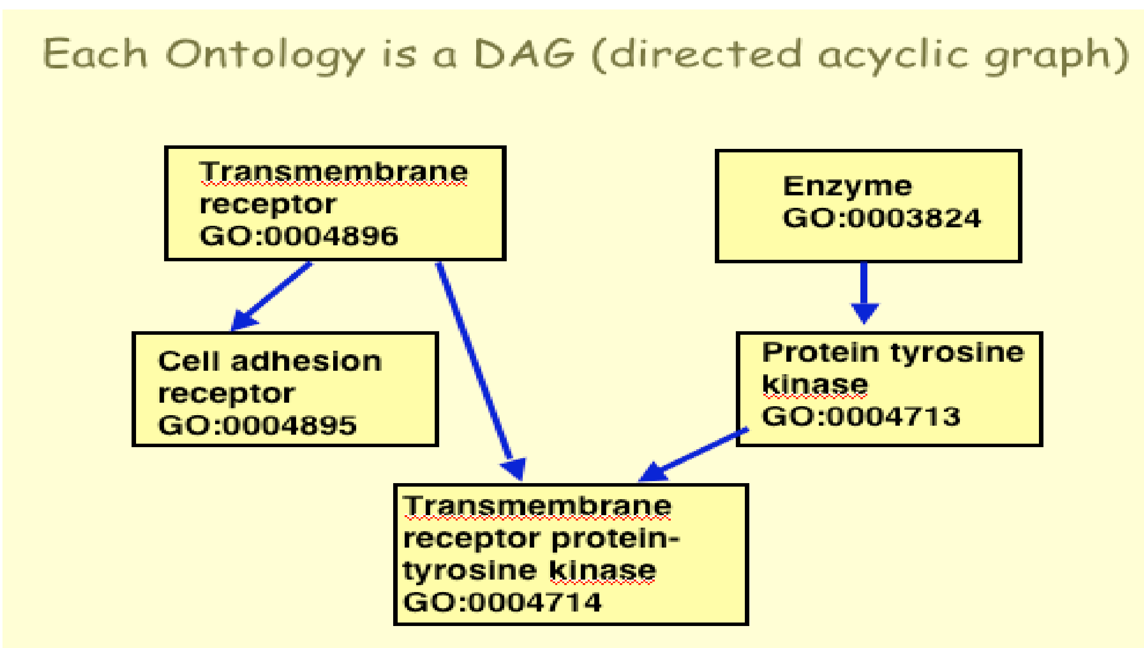
It is rather trivial to state that biological data are organised around specific phenomena: data can be broadly defined precisely as classified information about a process or object, which take different forms depending on how, where and by whom they are collected and expressed.[125] One of the few features (possibly the only one) that is common to all biological phenomena, and hence needs to be taken into account by any biologist no matter his or her specific expertise and interests, is their interconnectedness. Even the staunchest reductionist regards biological systems as complex ensembles of interdependent processes. The relations holding among those processes are crucial to the functioning of the system: disrupting one component of a biological system often means disrupting, and sometimes preventing altogether, the functioning of the whole. The complex relations among the components and processes characterising biological systems are the main focus of biologists' research. One important type of relation holding across biological phenomena is *mereological*: biological systems can be investigated via partitioning, that is, by locating some components of the system and studying the relations between those parts and the whole.[126] Another important type of relation is the one of *dependence*: the existence and functioning of any biological phenomenon is linked to the existence and functioning of a number of other phenomena. A given phenomenon can depend on others in different ways (ranging from the causal to the phylogenetic or derivational) and to different degrees: in some cases, two phenomena are so strictly dependent on each other that if one ceases to exist, the other one does too (as in the case of breathing and blood circulation in mammals).

These basic insights inform all functional and mechanistic explanations in biology, which are indeed grounded on the individuation of the mereological and dependence relations relevant to explaining a specific phenomenon.[127] Research carried out on Arabidopsis shares this aim. The common purpose underlying the generation of data about this plant consists in illustrating how different elements of Arabidopsis biology interact with each other, so as to explain its morphology, development and behaviour: this involves, for instance, looking at how genes group into clusters and contribute to phenotypic

[125] I thus take data to include anything from observation statements to various types of measurements. Remarks on what is meant for 'data' have been offered in section 5.1.3.
[126] Philosophers Wimsatt (1972), Cummings (1983) and Winther (2003, 2005a, forthcoming) have reflected on the importance of partitioning in scientific epistemology (Winther refers to it as a 'style of science' which he calls 'compositional biology').
[127] For instance, a mechanical explanation of behaviour requires biologists to individuate components of the relevant system that, through their interaction, produce that behaviour.

development or how each biochemical pathway is constructed so as to enable the transport of nutrients (as in metabolism) and information (as in immunological responses). When thinking about how to order and represent data relevant to those interactions, TAIR researchers had an intuition that crucially informed their choice of DAGs as an efficient representational tool. Namely, they noted how the child-parent relations holding among the objects of Java programming are, in fact, structurally isomorphic to the representation of the relations linking biological phenomena with each other. The asymmetry characterising the formal relations in DAG can also be thought to mirror the asymmetry in the relations holding among biological phenomena. The object-directed approach characteristic of Java could thus be exploited to visualise current knowledge about biological phenomena through a network of interrelated objects. To do this, one only needs to assume each Java object to represent a specific phenomenon and specify the relations holding between that and other phenomena (also represented by objects; see figure 5.8).

Figure 5.8 - This figure clearly shows how the interrelations linking some biological phenomena (in this case, cellular components) with each other can be represented as a DAG in the object-oriented approach. The biological nature of the relations is here as yet unspecified – what counts is the structural isomorphism with the structure characterising DAGs.



To be sure, this intuition was not original. The idea of exploiting the potential parallel between biological phenomena and the ‘objects’ of Java programming for the purpose of building databases had been circulating among model organism researchers since the beginning of the 1980s.[128] In fact, all sizable model organism communities, including the ones centred around Drosophila, C. elegans, E. coli, rats, mice, yeast and slime molds, have been struggling with the same problems experienced by TAIR. First, there is the overwhelming quantity of data needing storage and ordering, coupled with ample disagreement among practicing researchers about the potential significance and interpretation of those same data. Second, there is the problem of selecting *relevant* data. As TAIR curators were quick to realise for themselves, a database on a specific model organism could not possibly contain all existing data about that organism: too many data sets are incomplete, partially overlapping, inconsistent or tied to isolated (and possibly not corroborated) experiments and unpopular theoretical approaches. In this context, choices about which data sets to pick within any database needed to be made and, most importantly, justified.[129] It seemed that TAIR, like other databases for model organisms, needed to establish criteria for selecting and ordering Arabidopsis data. This issue turned out to be strongly tied to the practical problem of how to actually store these data and make them searchable, as we have seen in the previous section. Several representatives of different communities recognised that these common problems might best be solved in collaboration with other model organism communities. They thus started a joint project, which they called 'Gene Ontology Consortium' and which TAIR joined in early 2000, aimed at producing common criteria (in fact, a full-blown conceptual framework) for the building of databases dedicated to model organisms (Harris et al, 2004).

The GO consortium set itself to exploring the structural parallelism between the links among phenomena traced and explained through scientific research and the object-to-object relations imposed by Java software. Prima facie, the object-directed approach provides an ideal solution to the problem of classifying and retrieving data about interconnected biological phenomena: it simply requires to assume that every object used by the programme would represent a specific biological phenomenon, specify the relation holding among objects so as to represent the relation holding among the biological phenomena that the objects stand for, and categorise available data on the basis of their relevance to the phenomena thus represented. The resulting representation of interrelated objects is called *bio-ontology*: that is, a network of terms (such as 'gene', 'cell' and 'transport'), each of which denotes a specific biological object or phenomenon and is represented, as required by the object-directed approach, through its relations to other terms (denoting other objects or phenomena).[130]

---

[128] The initial idea can be traced to a collaboration between two Drosophila specialists, Suzanna Lewis (Berkeley Drosophila Genome Project) and Michael Ashburner (FlyBase, Cambridge), who started to think of IT strategies for storing, and elaborating on, Drosophila information since the early 1980s.

[129] TAIR insistence on the necessity of consistent criteria for the selection and ordering of data is precisely what makes this resource so different from all other existing databases for Arabidopsis research. The database maintained by the Nottingham Arabidopsis Stock Centre, for instance, is constructed on the principle that all data should be included, even if inconsistent, incompatible or even contradictory. This results into a much less structured database that is independent from a specific theoretical framework, but whose contents are much less difficult to read and interpret for someone who is not already steeped in the type of research and methods used by the researchers who originally contributed results.

[130] Note that the term 'ontology' to define a data set stems from research in information technology and is not related, at least in its origins and purpose, to the meaning attached to it within philosophy. A study of the relation between philosophical and IT notions of ontology constitutes an interesting project, which is now carried out principally by Barry Smith and his associates (2005). Within this dissertation I do not

Given the simplicity and representational efficiency of this framework for the classification and organisation of biological data, it is easy to understand why TAIR curators enthusiastically backed the project and incorporated bio-ontologies into TAIR. Still, the actual realisation of databases on the basis of GO network presents several conceptual and practical difficulties, which become apparent as soon as one examines a detailed definition of bio-ontologies. A recent review by Bard and Rhee (2004, 213) describes bio-ontologies as follows:

> an ontology is a formal way of representing knowledge in which concepts are described both by their meaning and their relationship to each other. Unique identifiers that are associated with each concept in biological ontologies can be used for linking to and querying molecular databases.

This definition points to three important characteristics of bio-ontologies, which considerably complicate the apparently straightforward task of representing biological phenomena as Java objects and which I shall therefore consider in turn: (1) the need for unique identifiers for each of the concepts[131] used; (2) the nature of their relationship to each other; and (3) their meaning (or definition).

Researchers at the GO consortium are aware that there cannot be a unique system of categorisation of biological phenomena, because the same phenomenon can be described differently depending on the scientific interests of the researchers studying it. Indeed, as remarked when discussing theoretical pluralism in Chapter 2, each epistemic culture in biology devises its own way to study and describe a biological phenomenon, which might or might not match the perspective offered by other cultures. For instance, the process of light absorption can be analysed in terms of the components of the plant involved or, in a different and yet complementary way, in terms of the underlying biochemical processes. On the other hand, the construction of bio-ontologies requires that there should be one and only one term (the so-called *unique identifier*) associated to each phenomenon. Terms acting as unique identifiers need to be precisely and unambiguously associated with a specific phenomenon (and thus, a specific set of data). Given the interdisciplinary context of GO users, in which the same term might be taken to refer to different phenomena depending on the epistemic culture in which it is used, the selection of unique identifiers becomes a daunting task. The GO consortium resolved this problem through a compromise between pluralism and the need for unique identifiers: three different bio-ontologies have been constructed, each of which concerns the same phenomena, but uses terms associated to different perspectives on those phenomena. One bio-ontology focuses on terms describing *biological processes* (e.g. aging, cell differentiation); another includes terms relating to *molecular functions* (such as transcription); and the third encompasses terms related to *cellular components* (e.g. nucleus, organelles, genes). This

---

concern myself with this issue: I wish to discuss the epistemological function that such 'ontologies' are acquiring within biology.

[131] Note that within my analysis I shall not differentiate between reference to GO objects as 'concepts' or 'terms'. This is because I am interested in what those terms stand for, rather than on their intrinsic epistemological status. For a discussion of the epistemological consequences of thinking of those terms as 'concepts', see Smith (2005).

ordering of information ensures that the data sets contained in each ontology are accurately matched to an appropriate term (and thus, a relevant phenomenon). By consulting all three bio-ontologies, users are able to check for themselves the difference between describing the same phenomenon as biological process, molecular function or cellular component.

The second issue in creating bio-ontologies concerns the nature of the *relations* that are to be represented. Of course, the unspecified child-parent relations holding among Java objects are not specific enough to represent the different types of relations holding among biological phenomena. Thus, the consortium chose to diversify the relations holding among objects into two categories: 'is_a' and 'part_of'. The first category denotes relations of identity, as in 'the nuclear membrane is a membrane'; the second category denotes mereological relations, such as 'the membrane is part of the cell'. Occasionally, a third category 'develops_from' is used to signal dependence relations, as in 'protein develops from amino acids'. [132]

Figure 5.9 - An illustration of two types of relationships holding among GO terms (in this case, terms denoting cellular components): 'is_a' and 'part_of'.

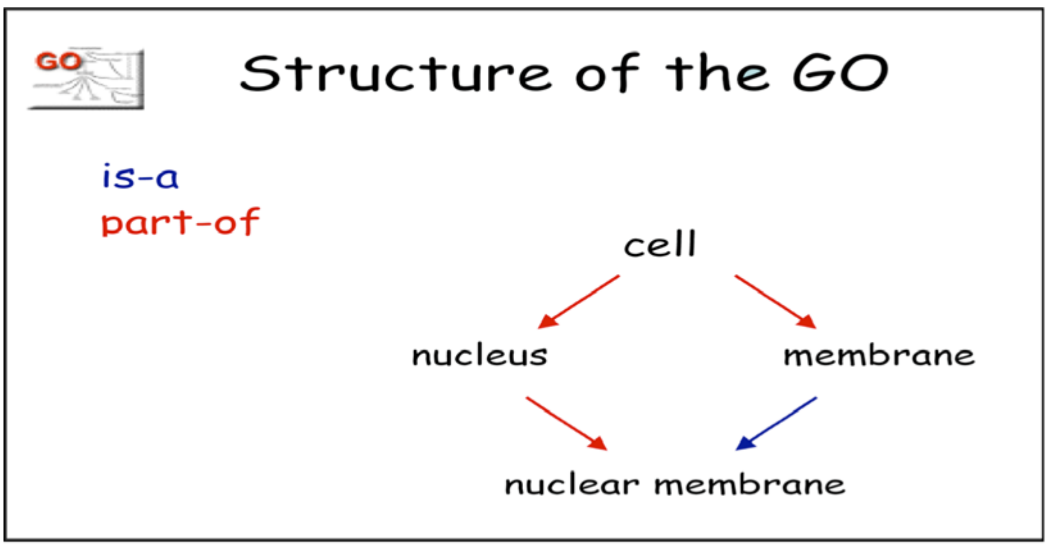
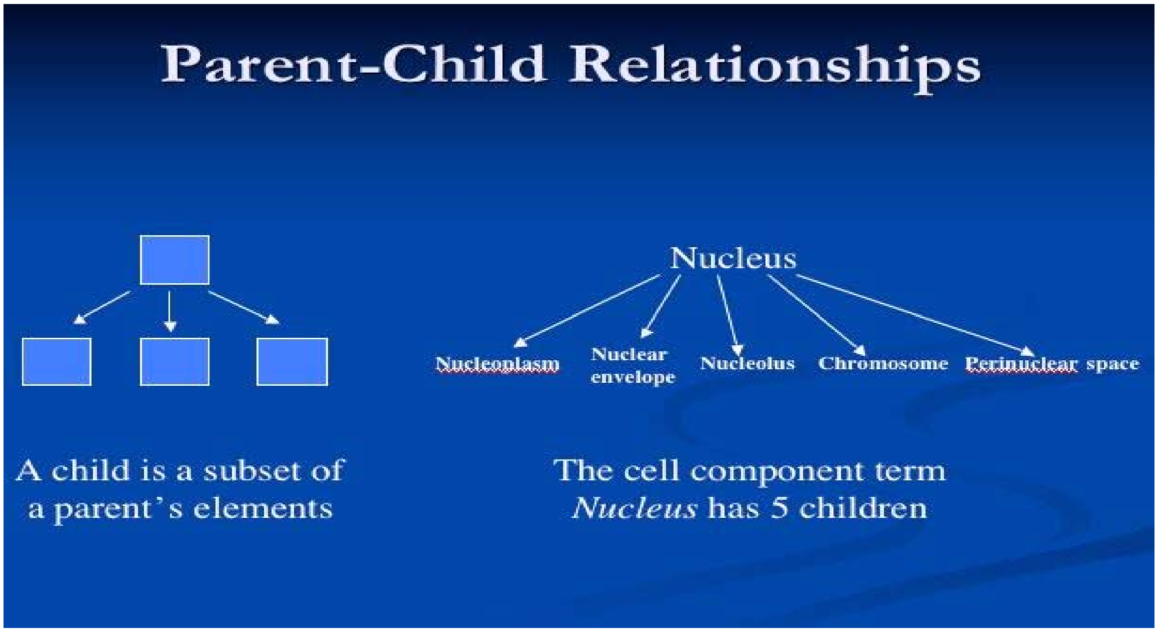


Figure 5.10 - Illustration of the parallelism between parent-child relationships in DAGs and in GO.

[132] Within other ontologies, for instance the ones employed to gather data about phenotypes, the categories of relations available are more numerous and complex: for instance, including 'measured_as' or 'of_a' relations.



The last problem emerging when constructing bio-ontologies is also the most complex. It concerns the necessity to assign a precise definition (a *meaning*) to the terms under which phenomena are classified and made to correspond with Java objects. GO concepts need to have a meaning, as the phenomenon that any one concept is chosen to denote needs to be defined in a precise way. Finding a precise definition of a biological phenomenon that would satisfy all biologists researching it is, however, a daunting task. It requires interdisciplinary consultations and agreement on a common vocabulary and common definitions, to be recognised and understood in the same way in all biological fields involved. Such agreement could not be reached purely through discussions among members of the consortium, since the expertise of the members does not voice all, or even most disciplines affected by GO definitions (for instance, developmental biology and genetics are overrepresented, while there is almost no input from biochemistry and ecology). Moreover, differences across biological subfields and epistemic communities do not concern only the vocabulary used to refer to the same phenomenon: even the definition of phenomena changes depending on the interests, instruments and goals of each discipline. For instance, the definition of 'pathogen' accepted by immunologists differs in scope from the one used by ecologists: the former indicates infectious agents who cause disease in their hosts; the latter encompasses all agents who interact with their hosts' biological functions, including the ones with beneficial, rather than damaging, influence (like in the case of symbiotic interactions).

A further complication with building agreement on the definition of GO terms consists in the very limited collaboration characterising communities working with different model organisms, especially when it comes to ways of classifying and interpreting evidence. Because of the 'founders effect' (see section 3.1) and of the amount of embodied and organism-specific knowledge involved, most scientists tend to work with one or two 'favourite' organisms and not to concern themselves with instruments, databases or even results produced by other model organism communities. This means that the terms and approaches used to investigate a specific phenomenon vary even across groups that possess the same expertise, but that are studying different organisms. The GO consortium has the merit of having been one of the first institutionalised platform encouraging collaboration and comparative work across model organism communities. This also means that there are many conceptual and material bridges to be built in order for collaboration to work.

The initial solution to this bundle of problems was to invite various experts (called, remarkably, ‘interest groups’) to act as consultants on different biological topics (and, thus, different sets of GO terms and relations to be defined). This initiative was later formalised into so-called ‘GO Content Meetings’. Organised every few months, these meetings aim at elaborating definitions for terms that prove particularly difficult to capture in a satisfactory, interdisciplinary way (such as ‘transport’, ‘enzymes’, ‘meiosis’, ‘metabolism’ and ‘pathogen’). Attendance to meetings varies depending on the topics to be discussed and includes experts from different disciplines, scientists from the GO consortium and researchers working on various model organism databases (including TAIR, whose curators play a prominent role in the organisation of these gatherings). The role of the invited experts is particularly important, not only in order to obtain feedback about what should and should not be included in the definitions, but also to verify how specialists from different fields might react to general definitions such as the ones elaborated in bio-ontologies.

5.2.3 Empirical Content: Data Mining and the Organisation of Evidence

After considering the conceptual difficulties of selecting appropriate definitions for the terms and relations used within bio-ontologies, this section looks at the last cluster of tools required in the making of TAIR. These are the tools employed by TAIR curators to gather and order data in the framework provided by bio-ontologies. Implementing the GO framework in TAIR involves the challenging task of locating appropriate sources of data to be inserted in the resource, as well as devising ways to display information about those sources. TAIR curators are extremely alert to the importance of locating good sources and, above all, providing users with details of who produced the findings retrieved through TAIR, where and how.

The first step taken by TAIR curators to gather appropriate data was to construct a search tool, called ‘PubSearch’, that automatically sifts through published literature about Arabidopsis research (as available in the main plant biology journals) to find the most updated and relevant information about specific Arabidopsis genes. Thanks to this tool, it is now relatively easy for curators to acquire an overview of what has been published, where and when about each Arabidopsis gene – and thus, to make a selection of the most relevant data. Further, TAIR curators want to incorporate knowledge embedded in notebooks, textbooks or academic memoirs by individual researchers: that is, knowledge that has not been published, but that is essential for users of TAIR in order to reconstruct the conditions under which data have been acquired. TAIR curators are alert to the importance, for users, to trace where each set of data was produced, how and by whom. It is understood that this is not just relevant in order for users to assess the significance of the data: it also allows them to reproduce those results, contact the researchers who originally produced it and, eventually, further their work or replace it with better results. Along with the data, TAIR curators want users to be able to retrieve information about (i) the methods used to obtain the data reported; (ii) the experimental set-up required; (iii) the original investigators, including their contact details if any further clarification or verification is needed.

Thus, TAIR aims at providing access to experimental processes and practices, rather than simply the resulting ensemble of data. This type of information is sometimes gathered by direct communication between curators and researchers: while it has the disadvantage of not being refereed[133], its scientific value is enormous, as it expresses some of the embodied knowledge necessary to the production of the results displayed in TAIR. Information about data sources is usually displayed along with the data themselves, as a result of each user search. Curators assign an *evidence code* to every data set, which clearly indicates the circumstances in which the data set was collected (table 5.1). For instance, the code IEA stands for Inferred from Electronic Annotation and it indicates data gathered from a digital source; IPI, or Inferred from Physical Interaction, refers to data resulting from the researcher's material intervention on the phenomena under scrutiny.

Table 5.1 - Classification of types of evidence and corresponding evidence code (note that evidence codes are not mutually exclusive).

Evidence Code Abbreviation		Evidence Code Definition
Computational:	IEA	Inferred from Electronic Annotation
Manual:	IDA	Inferred from Direct Assay
	IMP	Inferred from Mutant Phenotype
	IEP	Inferred from Expression Patterns
	ISS	Inferred from Sequence Similarity
	IGI	Inferred from Genetic Interaction
	IPI	Inferred from Physical Interaction
	TAS	Traceable Author Statement
	NAS	Non-traceable Author Statement
	ND	No biological Data Available
	IC	Inferred by Curator

[133] The lack of good quality standards for information provided on the data is a weakness of the project. For now, TAIR curators rely on peer review procedures by the journals issuing the papers that they use as reference for experiments and results; and on the reputation and record of researchers submitting their own experimental protocols. The hope is that, once the Arabidopsis community becomes familiar with the possibilities offered by TAIR, experimentalists themselves will start contributing to TAIR by bringing in new data & protocols as well as criticism of previously published work.

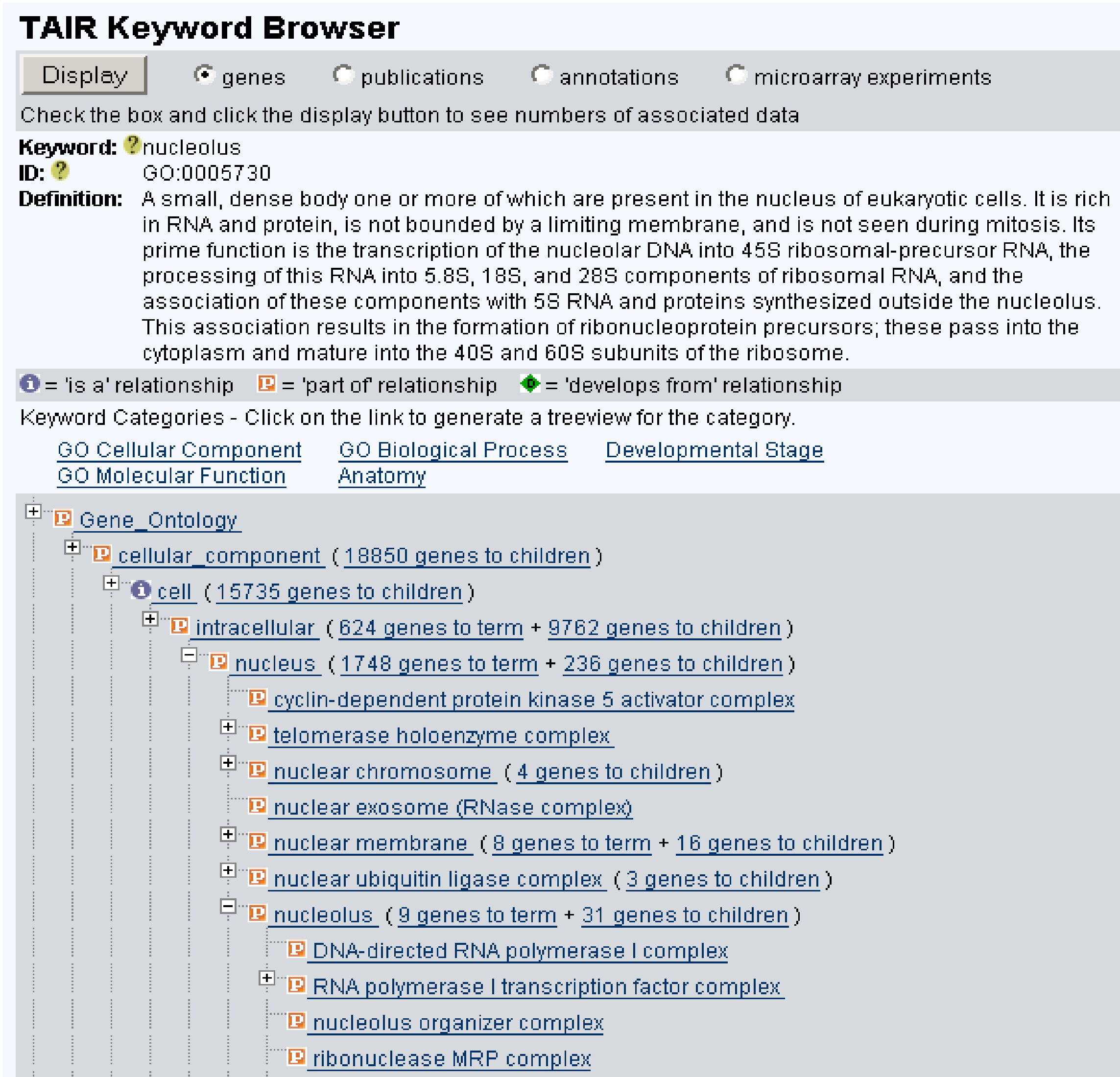
Apart from the evidence code, curators make sure that users can easily retrieve information about the experimental protocols, procedures and even personal annotations provided by the researchers who collected the data in the first place. This is supposed to guarantee that users are able to reproduce the results submitted to TAIR, or at least to trace the exact actions and experimental set-up used in order to generate them. A detailed list of the types of information involved can be found in table 5.2, which TAIR curators use when presenting their work to peers and prospective users.

Table 5.2. List of the features of experiments producing Arabidopsis data that curators would like to be retrievable directly from the TAIR website.

Experimental Information - Details	
Name	RNA Source
Description	Hybridisation protocol
Class	Samples: Name, Label, Description
Type	Number of Biological Replicates
Technology	Number of technical replicates
Plant Material	Normalisation
Anatomy	Image Analysis
Treatments	Contact
Growth Conditions	Publications

Once it is made clear which data should be integrated into TAIR (and with which additional information), curators proceed to collaborate with programmers so as to develop the actual databases. Data are collected and organised into data types as relevant to a specific GO term. The relationship among those data types is also adapted to the format required by bio-ontologies. This process is known as *GO annotation* and it results in a statement expressing the definition of a specific term, its relations to other terms and all the data relevant to both the term and the associated phenomena. An example of this annotation for the term ‘nucleolus’ is given in figure 5.11 below.

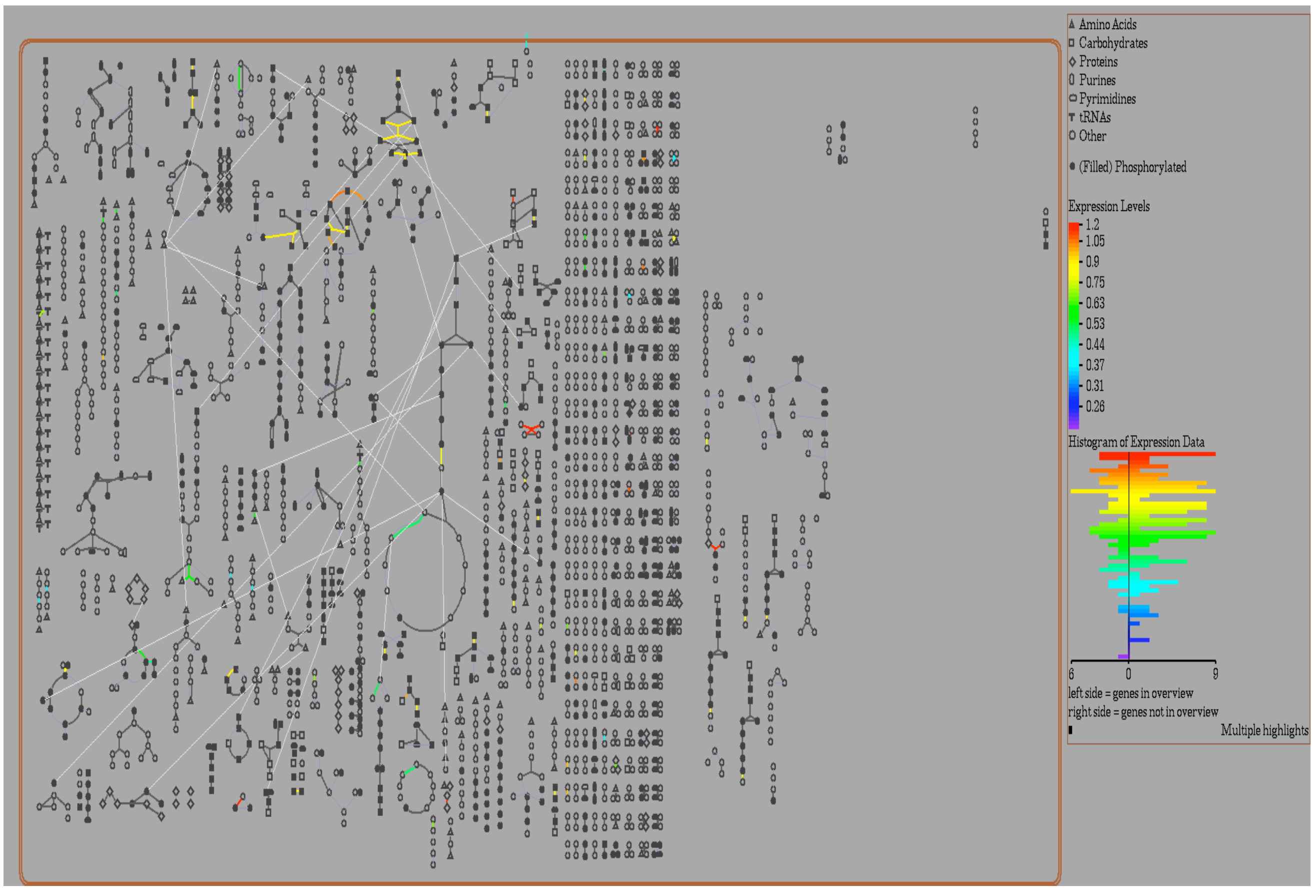
Figure 5.11 - Example of GO annotation of the cellular component 'nucleolus'.



The last step after annotation is the transformation of the data thus assembled and ordered into standard formats that can be fitted into Java programming, thus producing the visualisation of data accessed by the users. These standard formats, in which DAG structures are operationally fitted to the GO annotation, thus incorporating all data appropriately categorised, are referred to as *schemas*. Schemas are usually extremely complex, as they have to incorporate a lot of information (sometimes running into the hundreds of data). The production of schemas enables programmers to realise, insofar as possible, the vision initially proposed by TAIR curators, thus creating the visualisations of data that become available to the users (e.g. figure 5.12).

Figure 5.12 – *Arabidopsis metabolic pathways as displayed by AraCyc (the pathways themselves are represented by the white dotted lines connecting various components). It*

provides information about the components relevant to Arabidopsis metabolism at the cellular level, as well as about the expression level of genes controlling each of the components. One can click on any component (such as the triangles, representing amino acids, or the thin white line connecting several components and representing the pathway itself) to access information about the components themselves and the sources from which data have been acquired.



My reconstruction of the three main steps involved in the making of TAIR hopefully clarifies the difficulties involved in creating such a resource as well as the way in which it works. What remains implicit in my narration is the extent to which the personal judgement of curators, based on inferences from their own experience as researchers and potential TAIR users, influences the realisation of this resource. In the case of design, curators are in charge both of providing the initial vision and of checking that the programmers' realisation fits their expectations. When it comes to bio-ontologies, it is curators who gather with other experts during GO Content Meetings in order to evaluate definitions for the terms employed; and again, in order to gather evidence, curators determine the criteria for selecting data as well as for accounting for their sources. In short, the current format and functioning of TAIR is largely tied to the curators' judgement on how such a resource should work and what it should include. In the next sections, I focus on two types of knowledge used by curators to build TAIR. I intend to show that both of them need to be at least partly shared by TAIR users, who otherwise would not be able to interpret the information available towards improving their understanding of Arabidopsis biology. First, I look at the theoretical knowledge contained in bio-ontologies, and specifically in the GO network adopted by TAIR. Second, I examine the embodied knowledge guiding the curation (that is, the actual compiling of data by curators) of TAIR data sets and annotations.

5.3 Theoretical Knowledge at TAIR: GO and the Gene-Centric Perspective

There is, it seems to us,
At best, only a limited value
In the knowledge derived from experience.
The knowledge imposes a pattern, and falsifies,
For the pattern is new in every moment
And every moment is a new and shocking
Valuation of all we have been
T. S. Eliot, Four Quartets

5.3.1 GO: From Standard to Theory

In every phase of TAIR construction, curators bestow great attention to the prospective users of the resource. As I already remarked, TAIR is principally conceived of as a service to users. This is also true of one of its most important components, that is GO: in fact, TAIR curators present GO as a standardisation tool, that is, in the definition provided by Berg (Berg et al, 2004), as a 'coordination device' facilitating interdisciplinary research. To conceive of GO as a standard means to emphasise greatly its usefulness to users of TAIR, as well as the compatibility of GO to any type of research that users might be engaged in (hence, for instance, the use of different systems to capture data about processes, functions and components). The terms and relations specified in GO allow researchers working in different epistemic cultures to communicate with each other and discuss issues, methods and data of common interest. In this sense, GO terms constitute powerful standards around which research on model organisms can be organised. They offer a way to bring findings produced via diverse approaches, goals and tools under a unique framework, thus facilitating the exchange and consultation of such findings as an integrated whole.

Standardisation can, of course, be a double-edged sword. In their renowned study of classification, Bowker and Star (1999) claim that any standardised classification system is bound to mask, distort and/or disregard some of the characteristics of the objects that it includes as well as the specific interests, methods and goals of its prospective users. It is this homogenising effect that makes such classification systems so effective in providing information. By transcending disciplinary boundaries and local motives, classified data are made accessible to all interested groups; on the other hand, the process by which objects are selected and described in order to fit standardised classifications is often difficult to retrieve and assess for the users of such systems. Users therefore tend to rely on the available data without being aware of which information was left out of the classification system. Data that would not fit the given standards could still be relevant to the users' own concerns and research objectives. Standard classification systems can be as helpful as they are potentially misleading: an element of interpretation is unavoidable in order to build such systems, but users should be careful that the resulting homogenising effect does not constrain too strongly their thinking and reading of information therein contained.

The homogenisation of local knowledge into a 'global', standard framework is evidently also a characteristic of the GO project, as the GO consortium explicitly recognises. As a response to the possibility that GO might actually mislead its users, for instance by attributing incomplete, incorrect or incoherent definitions to its key terms, the creators of GO emphasise that the intent guiding their efforts is not to unify the whole of biology under the vocabulary and definitions proposed by GO. Rather, GO is supposed to act as a complement to existing resources and theoretical approaches in biology: it aims to provide an alternative vocabulary geared towards inter-community communication, which would improve the chances to exchange knowledge and constructive critiques among different subfields. Hence, GO terms are not intended as a substitute to the vocabulary used in any specific discipline: they are to be related, in ways determined by the relevant scientists, to the specialised terminology in use within each biological subfield. This is why TAIR curators are so careful in reporting sources of evidence and information on how to use TAIR on their website: the idea is to help users to 'unmask' the choices necessarily made by TAIR curators on the basis of GO standards, thus allowing users to challenge or refine their searches by examining the original sources for TAIR material (and, as a consequence, enabling users of different backgrounds to review each other's methods and protocols).

TAIR curators' strategy is not without problems. In the next section, I shall discuss issues relating to the use of embodied knowledge in constructing the resource. Here I want to focus on issues relating to the use of GO as a standard, particularly the claim that GO does not per se constitute an autonomous source for knowledge, but, rather, that it complements research already being done. There is a sense in which GO transcends this role of standardisation tool and does indeed become an independent source for non-local, integrated knowledge. Consider this recent definition of a bio-ontology, as reported in a review by Rhee and collaborators: 'an ontology makes explicit knowledge that is usually diffusely embedded in notebooks, textbooks and journals or just held in academic memories, *and therefore represents a formalisation of the current state of a field*' (2004, 221; my emphasis). The implication drawn in the latter part of this statement represents an important development: it now seems that bio-ontologies are not simply tools enabling users to retrieve information, but actually incorporate and express knowledge that is usually dispersed across a variety of publications and research groups. When used in TAIR, GO can thus be viewed as a 'formalisation' of the biological knowledge currently available about molecular and cell biology in Arabidopsis. I wish to argue that this implies a shift of epistemic status for GO, which comes to constitute a scientific *theory* about Arabidopsis biology rather than a mere standardisation tool.

This is a significant departure from the initial aim of the GO project and has notable implications for the TAIR project as a whole. Let me clarify the nature of this shift from standard to theory. To this aim, I refer to the work of philosopher Mary Hesse. Her reflections on scientific theory-making, themselves inspired by Duhemian insights on underdetermination and theoretical pluralism, prove very useful in order to specify the features of GO that make it work as a scientific theory about biological phenomena. I am here particularly interested in three aspects of Hesse's work, which I now briefly describe.

(1) Hesse's account of scientific theorising is based on what she calls the *network model* of theoretical science (1974, 4). In this model, theories are defined as networks of interrelated concepts. The meaning of each concept depends on the phenomena to which the concept applies (for example, the concept 'evolution' could be defined as indicating the process by which new generations of organisms are likely to acquire traits that allow them to survive better in their environment). Further, the meaning of each concept also depends on other concepts by way of law-like statements (for instance, Newtonian mechanics includes the concepts 'mass' of an object and 'force' applied to an object, which are related by the law-like statement 'force is proportional to mass times acceleration'). Theories can thus be propositionally expressed (by enunciating the series of law-like statements relating the concepts used).

(2) In such a network, there is no fundamental distinction between observational and theoretical languages. The choice, use and modification of each concept and relation expressed in the network depends on the empirical study of the phenomena that those concepts and relations are supposed to describe, and thus on the complex processes of experimentation, intervention and classification involved in empirical research (Hesse 1980, 84; 1974, 4-26). Hesse proposes that the theoretical representation of knowledge is instrumental to the goals and needs of empirical research (a good instance for this consists, in the case of GO, of the need to construct databases).

(3) Networks of concepts need to maintain an internal coherence and economy, depending again on the wishes and expectations of empirical researchers: 'without such organising conditions it is clear that a world described by even a small number of predicates in all apparently observed combinations is likely to become quickly unmanageable' (Hesse 1974, 52; see also 1980, 108). This suggestion can be interpreted in a number of ways. What interests me here is the idea that, in order to comply with the requirement of economy, concepts referring to the same phenomena should be reduced to a minimum in any one network; while, in order to comply with the requirement of coherence, the definitions of concepts and their relations should not contradict each other and should be expressed through the same format (that is, depending on the type of network at hand, the same vocabulary, imagery, formulation, or else).

I maintain that bio-ontologies, and specifically GO, abide to all three points and thus constitute a good candidate for a specific type of biological theory (among the several used in contemporary biology, as I illustrated in Chapter 2 when discussing theoretical pluralism). For a start, the GO network bears a striking resemblance to the structure of scientific theories envisaged by Hesse. GO indeed consists of a string of accurately defined, basic concepts with clearly outlined reference to specific phenomena. The relations among those concepts are extremely simplified, yet they are specified so that the whole network can be expressed and described as a series of law-like statements concerning the concepts in use, as in: 'nuclear membrane is part of cell', 'aging derives from cell differentiation' and 'mitosis is a cellular process'. Both structurally and semantically, GO concepts and relations do amount, according to Hesse's network model, to a theory about the relations holding among a number of biological phenomena. This is evident in GO creators' description of bio-ontologies as representations of biological knowledge: 'formal representations of areas of knowledge in which the essential terms are combined with structuring rules that describe the relationship between the terms. Knowledge that is structured in a bio-ontology can then be linked to the molecular databases' (Bard and Rhee 2004, 213).

The last part of this quote also exemplifies how GO satisfies Hesse's second point. The main motivation behind GO is to construct databases encompassing large amounts of data about all aspects of the biology of organisms. This means that the nature of the data available, as well as their format and the methods used by the laboratories that collected them, inform all decisions about which concepts to use, how to define them and relate them to each other. Concepts are thus not defined a priori, on the basis of purely theoretical considerations, but they are formulated to fit actual observations and, hence, the observational language used by empirical researchers. This became very clear to me when I attended the first GO Content Meeting[134], which was called to discuss the notions of metabolism and pathogeny, two concepts notoriously difficult to define and treated differently depending on the discipline of interest. Discussion among the immunologists, geneticists, developmental and molecular biologists present started from a tentative definition of the concepts and included mostly counter-examples to that definition, in the form of actual observations from the bench. For instance, the proposal that pathogens be treated as a independent category from organelles[135], popular among immunologists, was dismissed by ecologists and physiologists on the basis of specific cases of symbiosis, where pathogens turn out to be both symbionts and parasites of the same organism.[136] Here is how the argument goes: in these cases, pathogens cannot be treated as a separate, independent category from other microscopic components of the host's cell, since they also play a role towards the well-functioning of the cell as a whole. According to specialists in symbiosis and its role towards plant development, these pathogens should

---

[134] Held in Stanford, California, 28-29 August 2004.

[135] Organelles are secondary structures in the cell, such as mitochondria and rybosomes, which perform very specific functions, such as, in the case of mitochondria, that of converting organic material into energy usable by the cell in the form of ATP (through a process called oxidative phosphorylation).

[136] Certain kinds of bacteria (such as nitrogen-fixing bacteria or, as we have seen in Chapter 3, the Agrobacterium used to generate variants of Arabidopsis ecotypes) can have at the same time mutualistic and parasitic associations with the plants that host them. This generates a problem concerning the individuality of plants: are such bacteria components of the biology of their hosts, and thus part of the plants, or do they remain external agents? For more information about the conundrum generated by mutualistic, symbiotic and parasitic associations in biology, see Paracer and Ahmadjian 2000 and Tauber 1991).

therefore figure as 'part_of' the cell, rather than as a separate entity with no relations to it: thus, on the basis of a few observed cases, a whole theoretical category (the one of 'pathogen') is modified to fit a different context and definition. One of the keys to the successful functioning of GO is precisely the avoidance of distinctions between theoretical and observational language and related concerns.

The absence of a distinction between theoretical and observational language goes some way towards explaining the strong regulative role played by the requirements for coherence and economy in GO networks (Hesse's third point). As Hesse notes, coherence conditions are needed to manage the classification of predicates: 'not even the most rigorous inductivist would suggest that no further processing of the initial conditions should take place' (1974, 52). Coherence conditions in GO consist of a systematic structure, standard formats and limited relation types: these strictures are needed to avoid repetitions, internal contradictions and incommensurable definitions (which, if left alone to proliferate, would quickly render GO unmanageable). Another reason for insisting on coherence within GO stems from the obligation, acknowledged by the GO consortium, to serve one of the main goals of TAIR project - that is, the facilitation of the integration of knowledge among different biological fields. GO is a theory especially geared towards the integration of knowledge across different fields. This also requires defining GO terms and relations with an eye to the internal coherence and economy of the framework that GO comes to constitute.[137]

5.3.2 The Threat of Gene-Centrism

This last point is most important with regard to GO as a biological theory. This is because coherence among concept definitions and relations cannot be obtained without subscribing to a specific theoretical perspective, which provides the criteria according to which concepts have to be formulated and related to each other. James Griesemer characterises the goal of a theoretical perspective as 'coordinating models and phenomena' (2000, S348): in his view, according to which theories are themselves collections of models, a perspective is thus responsible for determining which aspects of

[137] Interestingly, some biologists I interviewed protested against the definition of GO as a biological theory: to them, the notion of theory is close to the one of 'hypothesis' – namely, biological theories express something we do not know yet, rather than what is already known. When I remarked that, within philosophy, the term 'theory' is used to indicate the expression of theoretical knowledge about phenomena, they all admitted that, indeed, GO is supposed to include 'all we know' about biology (and particularly plant biology, as within TAIR). The following passage (from an email sent to me on 27 August 2004 by Eva Huala as a follow-up on our interview) illustrates the point: 'I think that the TAIR database schema is not a pure representation of the relationships between biological concepts but is partly based on biology and partly utilitarian and based on technological considerations (to speed queries, avoid having to join too many tables in a query, etc.). Also the relationships we capture tend to be the more basic and universally accepted ones, not those that are currently evolving as biological knowledge is extended. Perhaps that's why biologists don't think of database design as the highest level of our field, analogous to a mathematical model for the universe. I think ontology structure is closer to a true conceptual model, but even there the relationships captured are not on the cutting edge of biology but those that are already generally agreed on and not controversial'.

a theory (and hence of the models constituting that theory) are relevant to phenomena, and how. Translated into my own terminology, this implies that theoretical perspectives unavoidably follow from the theoretical knowledge upheld in an epistemic culture: they signal the theoretical commitments made by adopting specific theories to describe and explain phenomena.

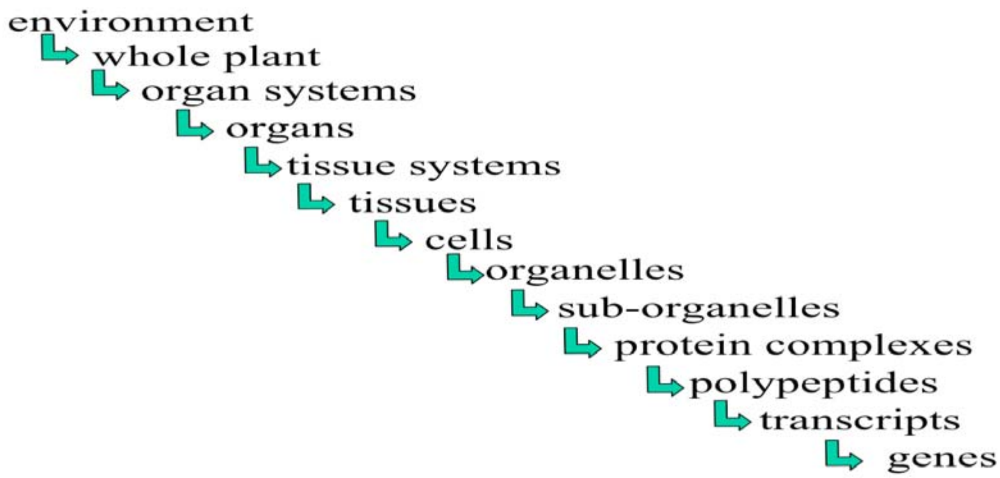
TAIR represents phenomena in Arabidopsis biology in a variety of different ways: that is, through the distinction, which I discussed in section 5.2.2, between bio-ontologies concerning biological processes, molecular functions or cellular components. These three types of bio-ontologies do not, however, amount to three different ways of interpreting biological phenomena: rather, they make use of three different classes of objects, which are however described by reference to the same overarching theoretical framework – that is, GO itself. The theoretical perspective upheld by GO, which I call *gene-centrism*, involves basing the whole bio-ontology on gene-related concepts (as exemplified by its actual name ‘ gene ontology’). In Rhee’s words, the annotation of data according to GO categories requires observable phenotypic traits to be ‘deconstructed’ into traits regulated by gene clusters (table 5.3).

Table 5.3. Classification of Arabidopsis data using multiple ontologies. While ‘primary’ and ‘qualifier’ ontologies differ (encompassing a component, a function and a process and correspondingly varied relations among terms), information is classified through association with a specific gene.

GENE	**Relationship**	**Primary Ontology**	**Context**	**Qualifier Ontologies**
AOC1	expressed in	Anatomy: leaf	1: during	Temporal: senescence
OST1	exhibits	Function: protein kinase activity	1: in 2: during	Anatomy: guard cell Process: response to drought
AG	is involved in	Process: specification of organ identity	1: of 2: in	Anatomy: petals Taxonomy: Arabidopsis thaliana

Initially adopted for purely pragmatic purposes, this commitment ends up informing both the observations and the concepts used in the network, with the result that the knowledge provided by GO is strongly biased in favour of gene-centric explanations of biological phenomena. For instance, the gene-centric perspective influences the mereological relations used to describe which GO elements are 'part of' others. GO describes mereological relations among phenomena via an extremely simplified, linear hierarchy going from environment to gene (figure 5.13).

Figure 5.13 - Biological organisation as visualised by Sue Rhee (presentation at the Carnegie Institute, Ca, Feb 6 2004).



Given that current research is often unable to specify the exact relationship holding between a gene cluster and the phenotypic traits associated to it, this system of classification will appear problematic to many researchers. Further, this representation has strong affiliations with the reductionist programme underlying much of molecular biology, according to which there is a direct and linear flow of information from genes to organisms (in keeping with the so-called 'central dogma' of classical genetics). This theoretical framework is in stark contrast with the perspective endorsed within fields such as evolutionary-developmental biology, or some strands of ecology, which stress the non-linear nature of exchanges of information among the components of a biological system.[138] Notably, many biologists interested in bio-ontology noticed the strong bias characterising GO and proposed alternative bio-ontologies, each of which exploits, more or less directly, a different theoretical perspective to describe and systematise data pertaining to a variety of aspects of organismal biology: there are, for instance, an 'Animal Natural History and Life History' ontology, a 'Pathway Ontology' and a 'Plant Attributes and Traits Ontology' (of which I shall say more in chapter 6). All these bio-ontologies, including GO, are grouped under the umbrella heading of Open Biomedical Ontologies.[139] Currently, much effort is being devoted to comparing and inter-relating concepts *across* these different networks. This constitutes an extremely interesting project in its own right, which would deserve a much deeper analysis than the few remarks contained in this thesis. For my purposes here, we need only keep in mind that GO is indeed part of a much broader and hitherto successful movement to order and systematise data sets on the basis of networks of concepts; and that each of these types of

---

[138] See Keller (2000) and Oyama (1985) on this.
[139] http://obo.sourceforge.net/.

bio-ontologies is obviously subscribing to specific theoretical perspectives and interests, thus constituting more than a mere translation tool for inter-community communication.

In fact, given the contentious status of the gene-centric perspective in biology at large, this bias represents a threat to the usefulness of GO as a standard: researchers working on a theoretical perspective that does not accept gene-centrism (such as, for instance, ecology or some strands in evolutionary biology) are reluctant to use databases constructed around GO. Worse still, some of these researchers might not be familiar with the theory around which TAIR databases are constructed, since their views on biological phenomena are informed by an altogether different theoretical perspective. This conditions their use of TAIR, since the organisation and display of TAIR data is dependent on GO and thus on the theoretical perspective therein expressed. In order to benefit from the information acquired as a result of a search through TAIR data, users need to interpret that information in their own vocabulary and in the context of their own research. This presupposes acquaintance with GO and an awareness of its theoretical commitments. Hostility to, or ignorance of, GO thus becomes a major obstacle both to accessing data displayed by reference to this framework and, most importantly, to evaluating the significance of those data towards understanding biological phenomena.

To sum up: the adoption of a gene-centric perspective is what makes the role of GO as a theory different from its role as a standard. Not only are curators forced to manipulate the available information in order to construct standardised formats through which information can be retrieved: they also subscribe to a specific theoretical perspective which guides their classificatory work, thus giving coherence to the concepts used. For the users of GO, this means that it is impossible to use these networks without implicitly subscribing to the theoretical perspective therein employed: a perspective that might be irreconcilable with their own, but which transforms GO into a coherent and economical source of knowledge about phenomena.

5.4 Embodied Knowledge at TAIR: Manual and Automatic Curation

> *Only the experimentalist knows the real strengths and weaknesses of any particular orchestration of machines, collaborators, interpretations, and judgements*
>
> *Peter Galison 1987, 244*

TAIR curators themselves acknowledge that it is impossible to construct such a database while remaining neutral over the interpretation of the data. They are well aware that all classification is at least partly normative: it imposes constraints on the interpretation of the material that is classified by way of the criteria adopted in order to select, order and display that material. In this sense, curators certainly recognise that the incorporation of a theoretical perspective in the making of GO (and thus, of TAIR) is unavoidable, as it is necessary in order to provide categories and parameters through which data can be retrieved. Reference to GO allows TAIR to clarify both the biological context to which data apply and the scientific context in which they are produced. Reliance on a theoretical perspective is, in this sense, an advantage rather than a necessary compromise.

Acknowledging the advantages of relying on a gene-centric perspective does not prevent curators from taking very seriously the problems encountered by users with different views of biological phenomena. Their response to this issue is to highlight that biologists need to learn to work with TAIR and that such learning involves mastering the search tools as well as being aware of the reasons why TAIR is constructed as it is. In other words, curators appeal to *transparency* and *user involvement* as the solutions to problems of accessing TAIR and interpreting data in the light of its conceptual framework. As long as the interpretive steps taken by curators while constructing the resource are documented and this documentation is made accessible to users, it is possible for users to evaluate those interpretive choices, their significance towards assessing the value of the data and, possibly, the mistakes therein. This implies that users should become active contributors to the TAIR project, rather than just passive recipients of its benefits. In other words, curators shift the responsibility for the interpretive choices made in TAIR to the whole Arabidopsis community: problems in the ontologies, they state, 'can be adequately handled if ontologies are felt to be owned by the field rather than just by the individual authors' (Bard and Rhee 2004, 221). Remarkably, this is not just a cunning excuse allowing curators to circumnavigate the nagging issues underlying the building of a multidisciplinary database. Curators do everything in their power to ensure that the reasons and constraints dictating their choices are indeed documented and available for their peers to consult and criticise. In fact, the TAIR website itself gives access to all existing reports on TAIR progress, to the minutes of each meeting among curators, to teaching materials (including animations on how to use search tools and GO) and even to all correspondence between curators and users (that is, curators' answers to complaints, requests for information and suggestions sent in by users).

This is an interesting way of engaging the tension between theoretical interpretation (allowing for integration) and accessibility to all potential users that I outlined in the previous section. The personal judgements that curators need to give in order to construct the resource are laid bare, so as to empower users to assess their work and decide whether to endorse those judgements.[140] I want to point to three problematic assumptions underlying this approach. One is the presupposition that users will indeed wish and/or have the time to engage with the extensive documentation mapping the curators' choice. This is debatable, as curators themselves are experiencing since launching the project: unless they already have a basic affinity with the perspective offered by TAIR, thus not really needing to be 'trained' and to read about the making of TAIR, biologists are too absorbed in their own research to actively engage with TAIR. As far as users are concerned, TAIR is a service provided to them so that they can access information with little effort: if it is not accessible with minimal background information, then it means that the service does not work. The second problem concerns the idea that curators can accommodate all incoming critiques, no matter how diverging from each other and from

[140] As I discuss in detail in Chapter 7, this approach makes TAIR a good locus for scrutinising the interaction of researchers located at the centre of Arabidopsis research (such as TAIR curators, all working in one of the most prestigious institutes in the world for this type of research) and researchers located at the periphery of the Arabidopsis community, that is, (geographically and/or intellectually) far from its central institutions.

curators' own ideas. This assumption is also clearly problematic, given precisely the argument (exploited by curators) that the construction of a resource like TAIR needs firm theoretical foundations (hence, arguably, endorsing one specific theoretical perspective). I shall come back to both of these important points in Chapter 7. The final problem lies in the assumption that the motivations and circumstances of the choices made by the curators can indeed be documented through a series of written texts, statements and illustrations. It is this assumption that I wish to discuss at length in the rest of this section.

In my view, it is impossible to fully report the background for the decisions made by curators by means of online resources (such as protocols, minutes of meetings and even step-by-step animations carrying users through the various processes and reasonings underlying the making of TAIR). To understand why this is the case, I turn to an exemplification of the problem, i.e. the work involved in compiling data into annotations appropriate for insertion into TAIR databases. This ensemble of tasks, which TAIR personnel refers to as *curation*, involves a series of decisions which, I argue, are motivated strictly by each curator's embodied knowledge of the phenomena and/or the type of data that he or she is annotating. Let me demonstrate how this works by briefly listing how curation is carried out at TAIR.

Each curator is assigned a set of data types of her or his own competence. Broadly, curators are divided up into sub-teams looking at three areas: functional adaptations, genomic micro arrays and metabolism. Given the prominence of genetic data in TAIR databases, the curator's job is currently focused on gene annotation, that is, the compiling of data relevant to each Arabidopsis gene to fit GO annotations and TAIR schemas. The curation of each gene can take between 20 minutes and four hours, depending on the type and quantity of data available. During that time, curators use PubSearch to look for the following three types of information: the first published paper on the cloning of the gene: that provides information on functional annotation; the most recent paper published on the gene, reporting the best information available about its biological function; and at least one paper giving information on location of the gene in the sequence. They then proceed to insert these pieces of information into the format required by TAIR, thus choosing:

(1) GO terms describing the biological function, activity, expression and location of the gene;
(2) GO relationship types, describing the relation between the gene and other aspects of Arabidopsis biology;
(3) An evidence code designating the type of evidence used (as documented in 5.2.3);
(4) An evidence description of the methods used to obtain the evidence (e.g. in situ hybridization);
(5) Date and attribution, that is, when and by whom the annotation was made (e.g. annotated on 12/15/2003 by TAIR curators)
(6) Reference: the papers used as primary source for the evidence (e.g. Huala et al., (1997) Science 278(5346):2120-3).[141]

[141] Material taken from a presentation given by Sue Rhee to Open Biological Ontologies Board on 6 February 2004.

Curators are convinced that the whole decision-making process underlying curation can be documented. In fact, they claim that the parameters through which data are selected and compiled are so straightforward, that the process of curation could soon be entirely automatised. For instance, the selection of the order in which genes are curated (the chronological priority of gene annotation) has been automatised in 2004, following the suggestion of an intern to the project. The automation works because it is possible to formulate an algorithm for the choice of genes, containing all the criteria necessary for prioritisation. In Rhee's view, increasing automation guarantees the possibility, for users, to easily trace the criteria used in selecting and compiling information. This would help access to TAIR by all interested users, no matter their own perspective and research interests.[142]

Despite all efforts at increasing automation, however, there are reasons why manual curation remains a crucial feature of curators' work. This is because the experience with experimental research accumulated by curators previous to their affiliation with TAIR is essential to the their understanding of the papers that they have to analyse in order to extract data for insertion into TAIR. The apparently simple information about a gene's location, expression and (especially) biological functions are actually very difficult to extract from a bunch of papers written by different authors for a variety of mixed purposes. The curator's ability is not so much to annotate the data so that they match GO terms and relations: it is to comprehend the experiments described in the papers, so as to be able to extract the relevant data (as well as relevant information about the sources of the evidence). All curators confirmed to me that, while in theory a researcher without experience 'at the bench' (i.e., experimenting with actual Arabidopsis plants) could curate genes, this has never been true in practice: all TAIR curators have experience as experimenters and they rely on it heavily in their interpretation of the data that they compile. Eva Huala, for instance, finds it difficult to compile information from areas and experimental approaches that she is not already familiar with. When selecting which data from the published literature are to be included in the database (and which ones should be discarded insofar as outdated or not representative), she needs to refer to her own knowledge about how to handle the relevant instruments, how to prepare and manipulate plants and cell cultures and how to read and compare results acquired through different experimental techniques. All of these abilities pertain to the broad category of knowledge that I characterised as 'embodied' in Chapter 2 – one of the features of such knowledge being that much of it cannot be expressed verbally or even pictorially, but is acquired through the exercise of specific (in this case, experimental) activities. Discarding evidence that is superfluous, dubious (because obtained through undocumented, non-replicable or even non-standard methods) or irrelevant to the curation of the gene or gene function under scrutiny is one of the main skills that TAIR curators need to exercise: only through appeal to such embodied knowledge can Huala assess the validity, accuracy and relevance of results obtained by other researchers.

A comparison among the elements that curators have at their disposal when annotating a specific set of data provides further clarification for how embodied knowledge operates in the curation of TAIR databases. Curators usually start from a specific item (a

[142] Interview on 11 August 2004.

component, process or function), whose identity is defined in relation to other items already in the database: for instance, a particular gene locus on a given chromosome. They use PubSearch to select publications that refer to that locus. This search may yield a small or large number of publications, depending on how much relevant experimental research has already been conducted. Now, the curators have to bridge a conspicuous gap between (A) the information available in and about the publications and (B) the information required by TAIR in order to classify the relevant evidence. (A) encompasses information such as the authors's names and affiliations, the way in which their research is classified in PubSearch (and other search engines) and the actual text of the publication. (B) includes a description of the evidence, a classification of the type of evidence displayed and of its relation to other pieces of evidence in related parts of TAIR, and eventual comments by the curator as to how the evidence should be interpreted and related to other pieces of evidence. Information of type (B) is often not directly displayed in the text of the relevant papers. Partly this is because authors are usually writing for an audience of specialists who do not need to have every detail of the adopted technique and preferred evidence spelled out. Further, as we have seen, it is because the evidence is classified according to standards other than the ones used in TAIR:

> 'the physical loci (genes resulting from the genome sequencing effort) are named according to their chromosome location. Chromosome-based names are a combination of organism name (AT), chromosomal location (1–5), G (for gene), and a unique accession number (e.g. AT2G34400). This nomenclature has been generally accepted as the standard nomenclature for *Arabidopsis* loci and replaces the previous one used during the sequencing phase, which was based on BAC names (e.g. F23H14.2)' (Garcia-Hernandez et al 2002, 243).

Curators are not just cutting information from the papers and pasting it into their databases. They need to interpret the content of the papers in the light of their own familiarity with the techniques and methods used in that field, so as to be able to extract the appropriate (B) type of information. In the case above, this means recognising the nomenclature used in research performed up to 2001 (BAC names) and, by adding information about its chromosomal location and accession number, transforming it into nomenclature acceptable for use in TAIR (e.g. AT2G34400).

The relevance of embodied knowledge to curating TAIR data implies that the longer a curator is unable to engage in experimental work, the more difficult it becomes for her or him to select and interpret data for annotation. Indeed, curators report that after some time spent as curators in TAIR, which implies a detachment from experimental work (only Rhee is still involved in experimentation at the same time), they start to lose touch with what happens at the bench. This is why, with the exception of the directors who guarantee continuity to the project, TAIR personnel shifts rather frequently: long-term curators need to refer to newcomers' fresh 'feel' for newly developed experimental practices in order to continue their work. As I shall discuss in detail in the next chapter, a physical interaction with the plants and the experience of experimental research seem to be crucial in securing individuals' ability of making sense of the data available so as to interpret them and arrange them in the database. Curation work, as much as the work involved in actually creating TAIR databases, thus involves reference to embodied knowledge that cannot be documented via reports, statements or illustrations on the website. Curators' decisions concerning visions, bio-ontologies, evidence and curation are largely based on their awareness of what it is like to experiment with Arabidopsis. Curators all come 'from the bench': they have been experimenters before entering the TAIR project, they plan to go back to experimenting as soon as their employment with TAIR is ended and they regret the impossibility to carry out experimental and TAIR work at the same time. This regret is not only due to their love for experimental research: all curators that I interviewed remarked how their contribution to TAIR is heavily informed by their awareness of what it takes to manipulate Arabidopsis in the laboratory. Again, this resonates with Hesse's analysis of theories: the expression of empirical observations as a network of interrelated terms (resulting in a series of law-like predicates, such as 'nuclear membrane is part of cell' in GO) necessarily implies a loss in non-verbalisable empirical information, such as the one that TAIR curators try to incorporate by their emphasis on sources of evidence (Hesse 1980, 93).

Arguably, this type of embodied knowledge is relevant to users as much as it is relevant to curators. That is to say, users need to appeal to both their embodied and their theoretical knowledge of phenomena, in order to assess and interpret the decisions underlying TAIR ordering and display of data. However, users cannot possess the embodied knowledge relevant to the interpretation of the data unless they have the same experimental experiences as TAIR curators. This is rather unlikely, given the richness of techniques and protocols available in Arabidopsis research. Even more troubling, in view of curators' claims, is the fact that such embodied knowledge cannot be reported on the TAIR website (and thus taught to TAIR users), since it is acquired via experiences of interactions with phenomena, rather than from verbal or pictorial representations of such interactions.[143] This disproves the claim that the motivations and background used in order to construct TAIR databases is (or could be) fully documented on the website. As a result, it becomes difficult for curators to appeal to transparency as a way out of the problem of understanding the database: even if users were trained according to TAIR guidelines, they still would not be able to trace and interpret all the circumstances and motives underlying the curators' judgement.

5.5 Using TAIR to Understand Arabidopsis Biology

The failure of the transparency argument given by curators lands the TAIR project in a peculiar situation. There is a tension between the interpreted nature of TAIR, made explicit by its reliance on the gene-centric theoretical perspective expressed through GO networks, and the role of TAIR as a depository of data of interest to the whole Arabidopsis community. Curators argue that this tension is unavoidable, given the need to select a specific theoretical framework in order to construct a resource of this scope

[143] This section has focused primarily on the embodied knowledge used by TAIR curators to construct the database. Yet, the fact that TAIR users also recur to their own embodied knowledge to interpret data visualised through the resource is crucial to my analysis and will take central stage in Chapter 6.

and size. Information, as well as frameworks for interpreting and using information, always comes from a specific context – which in this case consists in the epistemic culture endorsed by the TAIR group of curators and programmers. However, the embedding of TAIR into a specific theoretical perspective, as well as the use of GO as theories and extensive reference to the curators' embodied knowledge, risk to diminish the value of this project for its prospective users. What users fundamentally expect from TAIR, is that it facilitates their understanding of Arabidopsis biology: my discussion so far has highlighted how users, in order to use TAIR effectively to this aim, need to learn much of the theoretical and the embodied knowledge used in order to construct TAIR in the first place. I intend to dedicate this last section to a philosophical analysis of this situation in the light of its implications for understanding in biology.

As a starting point, consider again the definition of scientific understanding proposed in chapter 2: *the cognitive achievement realisable by individual scientists depending on their ability to coordinate theoretical and embodied knowledge that apply to a specific phenomenon*. My goal is now to start unravelling what this means and how such coordination of relevant theoretical and embodied knowledge should be carried out. My examination of TAIR construction and use allows me to clarify both of these aspects in the case of Arabidopsis research based on TAIR: I can now exemplify what theoretical and embodied knowledge amount to in this case and, most importantly, I can specify some *modalities* through which reference to these two types of knowledge may yield understanding. In fact, the basic point illustrated by my analysis of TAIR is that not anything goes in order to acquire understanding in biology. The required coordination of theoretical knowledge and embodied knowledge needs to be accomplished *skilfully*, that is, in ways appropriate both to the features of the phenomenon at hand and to the social and material conditions in which research is carried out. This section thus includes a discussion of what I call epistemic skills, that is, the skills expressed in the actions undertaken in order to acquire understanding of a phenomenon. Using TAIR to understand plant biology requires an extensive set of such skills: I distinguish between theoretical and performative skills, both of which are needed for users to be able to understand the significance of TAIR data to the biological phenomena of interest to them.

5.5.1 Introducing Epistemic skills

Prima facie, epistemic skills may be broadly defined as the abilities to carry out a number of activities in order to increase one's understanding of reality.[144] As a result of exercising those abilities, scientists become aware of how their experiences, as well as the concepts and tools that they use, fit or challenge available explanations of phenomena. Thus, as I shall illustrate in this last section, these abilities constitute indispensable conditions for the acquisition of understanding, precisely because they

[144] Giving a full account of the notion of epistemic skill is an ambitious task, which certainly deserves more attention than it has hitherto received from philosophers of science. My present reflections are meant to highlight some crucial aspects of this notion in the context of biological research, without pretending to exhaust its many dimensions and applications.

allow a fruitful coordination between the theoretical and embodied knowledge of phenomena.[145]

Epistemic skills may be partly innate, such as the skill of drawing an object (which depends at least to some degree on the talent of the individual attempting such action), yet they are most often acquired through the imitation of others and/or through experience, for instance by iterative trials. Skilful actions are, in a sense, instrumental: they are necessarily targeted towards the achievement of a specific goal, which in the case of epistemic skills is an improved understanding of (some aspects of) reality. At the same time, the notion of skill concerns principally the means and manner by which action is undertaken. In fact, an action can be judged as skilful even if it does not result in the accomplishment of its intended goal. This is because the successful achievement of a particular aim involves more factors than just the intentions and ability of the individual acting to that aim: adverse conditions in the environment, bad timing, social context, interference with other individuals' goals and actions – all these factors can influence the outcome of an action, no matter how skilfully that action is carried out. For instance, a Japanese sword master intending to slay his enemy might fail to successfully accomplish his goal, for reasons entirely beyond his control: for instance, in case the enemy carries a shotgun and shoots him while he is still too far to strike; or, in the admittedly unlucky case of lightning striking him while he is raising his sword to inflict the final blow. Failure due to factors beyond his control does not make his attack less skilful.

I take this to illustrate that the assessment of an action as skilful depends as much on the manner in which the action is undertaken as it depends on its effectiveness in achieving the intended goal. In other words, possessing an epistemic skill implies more than the ability to perform an action: it requires performing that action *well*. This means that we need criteria to determine what it means for an action to be well performed. However, these *regulative criteria* are extremely context-dependent and thus arguably impossible to classify in an analytic fashion. They are dictated by factors as disparate as the nature of the goal to be achieved; the interests of the person performing the action; the social as well as the material context in which the action is carried out; and, last but not least, the tools available in order to carry out the action. Consider as an example the building of a wall. Regulative criteria for the actions to be undertaken to build a wall skilfully include: geometrical precision, which is dictated by the nature of the goal to be achieved (by definition, the wall should stand strong and perpendicular to the ground); the aesthetic preferences of the person building the wall as well as the architectural conventions adopted by his or her social context (dictating the colour and shape of the wall); the degree of resistance expected from the wall, which depend on its position and exposure to the environment (a wall built in a region subject to storms, hurricanes or earthquakes will have to be much thicker and have deeper foundations than a wall built in a region with a mild climate; a wall built to separate two rooms will have different features from the

[145] Note that epistemic skills are used to acquire various types of understanding, including – but not limited to – scientific understanding. I do not intend to offer an a priori demarcation between epistemic skills needed to understand scientifically and epistemic skills needed to understand in other ways (for instance, in a political or spiritual way). As I discuss in chapters 2.3 and 7, such a demarcation can only be drawn a posteriori by looking at the actual practices adopted and sanctioned within scientific communities.

front wall of a house); and, finally, the time employed and the manner of disposing the bricks, which depend on the composition of the bricks themselves (and thus the humidity and insulation capacities of the material used) as well as on the tools available to the builder (type of trowel, cement, etc.).

What I would like to claim here is that the regulative criteria relevant to epistemic skills used in biological research are no less local and context-dependent than in the case of skills used in masonry. This extreme context-dependence makes it uninteresting, in my view, to try and list those criteria without looking at specific practices. Still, skilful actions in masonry seem to have two features in common with skilful actions in biological research (and possibly in other contexts): (1) they involve the exploitation of tools in a way deemed appropriate to effectively pursue a proposed goal; and (2) judgement on what constitutes an 'appropriate' course of action largely depends on standards upheld within the relevant social context. In the case of epistemic skills used to increase biological understanding, the tools to be exploited may range across models, theories, experimental instruments, features of the environment as well as samples of phenomena themselves, while the relevant social context is constituted by the community of scientific peers in charge of examining the methods and results of any research programme. The best way to capture these insights is, I propose, to adopt the following definition for an epistemic skill exercised in biology: *the ability to act in a way that is recognised by the relevant epistemic community as well suited to understanding a given phenomenon.*

5.5.2 Theoretical and Performative Skills

In the course of the next sections of this chapter, I shall present and discuss two types of epistemic skills. After having discussed their use in Arabidopsis research, I shall argue that, if used together, these epistemic skills indeed enable biologists to coordinate theoretical and embodied knowledge as required to understand a given phenomenon.

The first type of skill, which I refer to as *theoretical*, involves mastering the use of concepts, theories and abstract models – in short, being able to manipulate various expressions of theoretical knowledge - towards the understanding of a specific phenomenon. As I shall show through the example of TAIR, theoretical skills enable biologists to reason through given categories and classification systems according to specific inferential rules, while at the same time judging the validity of those categories and rules with reference to alternative theoretical frameworks applying to the same phenomena. The second type of skills encompasses the *performative skills* enabling biologists to exploit material resources towards the acquisition of biological understanding. In other words, performative skills consist of the ability to interact with the environment (including laboratory equipment and the plant specimens themselves, as specified in the next chapter) in ways relevant to the study of a specific phenomenon. Performative skills are different from theoretical skills insofar as they can only be acquired through interaction with phenomena. A biologist can be taught how to cultivate a plant so that it will develop as required by a specific experimental set-up. However, the corresponding performative skill is acquired purely through practice, that is, by trying over and over again to act in the desired manner, thus gradually adapting movements and sense-perception to the tools and materials used in an experiment, as well as to the standards enforced by the research community of interest.

Both theoretical and performative skills are expressions of embodied knowledge – that is, as I discussed in Chapter 2, knowledge about how to act and reason as required in order to intervene in the world, improve control over phenomena and handle representations of those phenomena. In what follows, I shall discuss features of these two classes of skills. Another important type of epistemic skill consists in the *social skills* denoting the ability of researchers to behave and express their insights in ways that are recognised by their peers or/and other participants in their social context. As I highlighted at the end of chapter 2, this social dimension is extremely important to the definition of understanding as 'scientific'. I shall therefore come back in detail to social skills in chapter 7, where I discuss the epistemic significance of the structure and organisation of the Arabidopsis community.

Let me start by exemplifying what counts as a theoretical and as a performative skill in the context of using TAIR to understand Arabidopsis biology. This can be easily exemplified by looking back at the making of TAIR, as documented in this chapter, and analysing which epistemic skills were used to produce TAIR visualisations of data. TAIR curators employed a number of theoretical skills to create the resource. Most notable among those are the skills required to create GO and use it as a framework for data-management in TAIR. Those include, as we have seen, the ability to associate sets of data to specific concepts on the basis of the definition of those concepts provided by GO; the ability to link concepts by means of simple relations of the type 'is_a' and 'part_of'; and the ability to interpret and present the data of interest in terms of the gene-centric perspective characteristic of GO. The performative skills required to handle TAIR can be divided in two categories. The first consists of the IT skills enabling biologists to access the data and exploit the virtual environment provided by the resource to play around with those data in accordance with their own research objectives; examples of these performative skills are running TAIR searches, order data according to parameters of interest, eliminate from the visualisation elements that are not relevant to the user's specific concerns. The second set of performative skills involved in the use of TAIR consists in the skills gained by interacting with the phenomena themselves (that is, the Arabidopsis plants), as discussed in section 5.4. These skills enable biologists to trace the empirical significance of data displayed in TAIR. For instance, TAIR indications about the sources of evidence for a specific data set are only useful to a researcher who is well versed in similar research methods, knows how they can be performed and what that implies for the applicability of the data obtained through them. As I shall describe in more detail in Chapter 6, relevant performative skills in this case are the ability to manipulate plants and to experiment on them through a variety of techniques; the ability to identify the empirical evidence for and content of TAIR; to store, sow and germinate seed; to work with Java object-directed programming (IT skills). The possession of such performative skills makes researchers aware of the relation between evidence, as reported in TAIR, and the phenomena from which it is obtained.

5.5.3 Combining Skills: Many Types of Understanding

As it is evident from the above list, some theoretical and performative skills are prerequisites even just to access TAIR databases, let alone to use them effectively in the context of one's own research project. TAIR users certainly need to be able to structure descriptions of biological processes and data through the parent-child relationships characterising DAGs. This epistemic skill is crucial, as this is the basic structure used in GO to relate concepts with each other. As I showed in section 5.3.2, TAIR users might also profit from the ability to recognise the theoretical commitments and motivations underlying GO. To acknowledge the role of GO as a theory about Arabidopsis biology makes it easier to critically assess the validity of GO definitions and relations, consider possible alternatives and evaluate how significantly the gene-centric perspective affects the interpretation of data that is embodied in TAIR visualisations. Further, biologists need to possess the appropriate IT skills to access TAIR and exploit its features.

The unavoidable necessity to exercise specific theoretical and performative skills in order to consult TAIR constitutes a problem for TAIR users, most of whom do not possess these skills upon accessing the resource. As I already illustrated, TAIR curators are aware of the importance of epistemic skills to use TAIR. The TAIR website thus provides as much guidance as possible on how to acquire such skills: through simulations of the functioning of TAIR searches, as well as information on the origin and issues surrounding GO and DAGs, TAIR curators hope to provide users with enough theoretical and IT skills to handle and profit from TAIR visualisations of data sets. It remains to be seen whether the information and training given on the TAIR website is enough for users to acquire those skills. TAIR curator Leonore Reiser, who is responsible for education and outreach to the larger biological community, maintains that biologists should follow courses on TAIR as part of their academic training, so as to ensure that they possess such essential skills.[146] In the absence of such training, now unavailable in the majority of biology departments, it is difficult for TAIR to be as accessible to users as it is intended to be: the acquisition of the right skills is a necessary condition for TAIR to fulfil its function as a service to the Arabidopsis community as a whole. TAIR cannot be seen as a useful tool towards enhancing biologists' understanding of Arabidopsis biology, unless adequate training and practice in the relevant theoretical and performative skills is enforced upon the members of the Arabidopsis research community.

Does this imply that TAIR users should possess precisely the same skills as TAIR curators in order to be able to use the resource to enhance their biological understanding? The answer to this question is an emphatic no. Surely, some of the curators' skills are simply indispensable to access and use the resource. However, the number of epistemic skills that users can choose to exercise in order to interpret and manipulate TAIR data is huge and depends largely on their own purposes, habits, beliefs and expertise. An obvious example is provided by the performative skills enabling biologists to carry out experiments on Arabidopsis plants. As I pointed out in section 5.4, the ability to actually interact with experimental materials is paramount to the curators' success in annotating

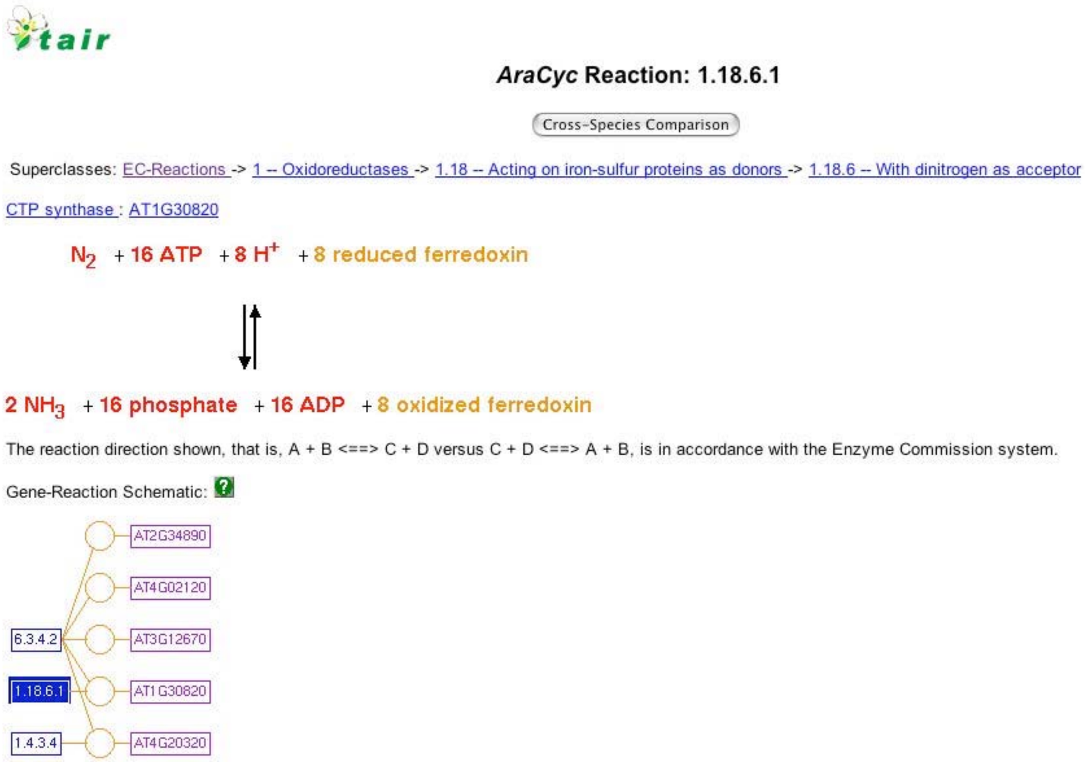
[146] Interview on 19 August 2004.

data. TAIR curators need to refer to their embodied knowledge of what it means to experiment on plants with one or the other technique, in order to assess which data are most relevant and useful for inclusion into TAIR. Equally, TAIR users need to refer to their performative skills in order to interpret the evidence, sources and methods reported by TAIR. However, each TAIR user possesses a (somewhat) different set of performative skills, depending on his or her laboratory experience, type of research, social context, disciplinary standards and even on their geographical location (as laboratories located in less financially powerful regions of the world might provide less access, if any, to the most recent instruments, materials and techniques for experimental research). Indications as to which of those numerous performative skills are most relevant to the interpretation of TAIR data are not to be found on the TAIR website. This is not only because it is practically impossible to convey that information through a virtual medium. It is also because TAIR curators do not value the extent to which their own performative skills impact their choice of tools and vision in visualising the data: while attempting to document the embodied knowledge of the researchers who provide the original data (as they show by reporting sources of evidence in the database), they do not take account of their own embodied knowledge as curators of the resource. The curators' lack of reflexivity with respect to the performative skills that they use to select and annotate data results in ambiguity as to the empirical value of the data themselves – a confusion that cannot be clarified or avoided by reference to TAIR alone. In fact, users are likely to possess entirely different sets of performative skills than TAIR curators, thus interpreting the empirical significance of TAIR data sets in a potentially different way than the one envisaged by TAIR curators.

Similarly, users do not need to learn precisely the same theoretical skills as employed by TAIR curators to access and use TAIR data for their own research purposes. Only some of the theoretical skills used by curators to build TAIR are essential to use the resource. Other theoretical skills used in the making of TAIR are only required if users wish to understand Arabidopsis biology *in the same way as TAIR curators do*. A biologist might refer to TAIR data without being aware of the theoretical perspective used to classify and present them, or even without being acquainted with GO. Further, possessing the theoretical skills relevant to reasoning through GO concepts is necessary to working through data in the way proposed by TAIR curators, but does not preclude using those data in other ways. Many biologists possess theoretical skills that are unknown to TAIR curators and which enable those biologists to interpret TAIR data in ways that curators would not conceive of (not even while 'playing the user', a practice which, as we saw, is restricted to the curators' own training and experiences). In the likely case that, as a result, their interpretation differs from the one supported by GO itself, they might draw very different conclusions from viewing those data than biologists endowed with another set of theoretical skills. Consider the following example.

A biologist accesses the resource (and specifically the search tool AraCyc) in order to find data concerning reactions in Arabidopsis molecular biology that involve nitrogen. As a result of her search, she finds the visualisation of data reproduced in figure 5.14.

Figure 5.14 – Result of the query ‘nitrogen’ under the search category ‘reactions’ in Arabidopsis metabolism.



A TAIR curator, well versed in the interpretation of these data according to the gene-centric perspective enforced by GO, would acknowledge that the query triggered a response by a sub-class of the term ‘reactions’ that is called ‘EC reactions’; within that class, the TAIR database selected a specific component of ‘oxydoreductases’, that is, the reactions ‘acting on iron-sulfur proteins as donors’. Now, a TAIR curator would not interpret this result as representing all the known reactions in Arabidopsis metabolism that involved nitrogen: rather, she would rephrase her query so as to fit the format of TAIR, by using a more specific term that is already present in the GO network of concepts. In playing around with the terms used for the search, she would immediately find out that the category ‘oxydoreductases’ alone actually contains 22 subcategories, of which at least half involve nitrogen-fixing reactions (such as, for instance, ‘acting on the CH-NH(2) group of donors’).

A user who is not acquainted with the GO network of concepts nor with its gene-centric organisation would, however, be puzzled by the result shown in figure 5.14. Surely, he would claim, there are many more reactions in Arabidopsis metabolism that involve nitrogen-fixing: as all plant biologist know well, nitrogen is one of the most important substances in the development of plants and its assimilation requires a number of complex molecular reactions. Instead of rephrasing the query in a way that is conducive to finding better-fitting results, he would conclude that TAIR does not contain the data that he needs; or, he would be deceived into thinking that the reaction portrayed in figure

5.14 is a particularly important nitrogen-fixing reaction and that he should study it in more detail. The example illustrates how a variation in theoretical skills with respect to TAIR curators determines a change in the understanding of nitrogen-fixing reactions acquired by the user thanks to TAIR.

I can now conclude that TAIR users do not need to possess precisely the same combination of theoretical and performative skills as curators, in order to use the resource for their own research purposes. They can use a variety of different epistemic skills to access and interpret the resource for their own research purposes. This has important epistemological consequences for the type of understanding of Arabidopsis biology that they acquire through consulting TAIR data sets. As I claimed in Chapter 2, biological understanding emerges from the skilful coordination of theoretical and embodied knowledge available about a phenomenon. In this chapter, I made clear how using different skills can lead to a different interpretation and handling of the same data and conceptual categories: the term ‘nitrogen’ and the data associated to it in TAIR need to be handled in ways similar to the ones employed by TAIR curators, in order for their empirical significance to be interpreted in the same way as is done by TAIR curators. In the case that different TAIR users employ different skills in accessing the same data, their interpretation – in fact, the very understanding of the phenomena in question that they provide – is likely to change. This means that biologists can obtain different understandings of the same phenomena in Arabidopsis biology, depending on the sets of theoretical and performative skills available to them in order to interpret and use the relevant data sets and conceptual categories that they find in TAIR.

This point illuminates an important aspect of biological understanding as a whole, to which I will come back in more detail in Chapters 6 and 8: that is, that understanding in biology varies not only in degree, but also in quality and substance. The possibility of a pluralism of epistemic skills, and thus of picking different combinations of theoretical and performative skills in order to obtain biological understanding, corresponds to a pluralism in the interpretations of data obtained through the exercise of those skills. As I shall discuss in more detail in chapter 8, there can be different types of understanding, depending on one’s skills, interests, commitments and research context.

This pluralism does not diminish the epistemological significance of understanding biological phenomena, but rather enhances the richness and the potential applicability of biological knowledge: as we shall see in Chapter 7, pluralism among ways to understand organisms plays an important role in biology, as emergence of and dialogue among alternative interpretations and ways of understanding is instrumental to the growth and refinement of scientific knowledge. However, the existence of various ways to reason about, interpret and understand biological phenomena constitute, at the same time, a potential obstacle to meaningful communication across different communities using TAIR. Researchers with different skills and backgrounds can certainly communicate with each other, but in order to do so they need to acquire either (a) some overlapping skills allowing them to understand phenomena in a similar way; or (b) an awareness of how different their understanding of phenomena actually is, as well as of the extent to which such difference is due not only to their theoretical perspective on the phenomena in question, but also to the skills that they employ in coordinating their theoretical and embodied knowledge.

The potential difficulties in communicating experienced by researchers with different skills and different understandings of plant biology makes TAIR's goal of integration more difficult to achieve than expected by TAIR curators. Does integration of data gathered in different biological subfields involve adopting a common outlook and common skills in order to understand biological phenomena? While TAIR curators (as well as the GO consortium) would like to answer in the negative, by pointing out how flexible TAIR is to the needs of users with differing interests and expertises, I hope to have demonstrated that at least some common skills and background knowledge are required in order to obtain integrated knowledge about a specific set of phenomena (such as, in the case of Arabidopsis, the structure, functioning and development of a plant). I do not think it possible to obtain a shared understanding of Arabidopsis biology as an integrated whole without adopting a specific theoretical perspective (in the case of TAIR, a gene-centric one) as a common platform on which to build the necessary tools and concepts. This does not necessarily represent a defeat to biologists searching for an integrative approach to organismal biology that is respectful of the pluralism in tools and perspectives characterising biological research as a whole. However, biologists using TAIR need to be well aware that the necessity to favour a gene-centric theoretical perspective over other perspectives is due to purely pragmatic reasons and should not be allowed to constrain the biological understanding acquired through access to the resource.[147]

[147] As illustrated in O'Malley and Dupré (2005), a similar debate is currently surrounding the various attempts to integrate biological data that call themselves 'system biology' (in fact, TAIR and especially the bio-ontologies project could themselves fit under this umbrella term). I shall come back to the links among integration, theoretical unification and understanding in biology in Chapter 8.

Chapter 6. Modeling Arabidopsis: Where Theory and Plants Meet

> *This capacity of a thing to reveal itself in unexpected ways in the future, I attribute to the fact that the thing observed is an aspect of reality, possessing a significance that is not exhausted by our conception of any single aspect of it. To trust that a thing we know is real is, in this sense, to feel that it has the independence and power for manifesting itself in yet unthought ways in the future*
>
> *Michael Polanyi 1958, 132*

The previous chapter focused on the theoretical framework adopted in TAIR and its significance for the biological understanding of Arabidopsis provided by that resource. In the conclusion, I pointed to the combined exercise of theoretical and performative skills as an indispensable condition towards acquiring such understanding. Further, I noted how, in order to make sense of the empirical significance of TAIR data, TAIR users need to possess performative skills (such as an awareness of how to interact with plants and acquaintance with some experimental techniques, sources and instruments) that cannot be acquired by connecting to the internet and downloading images. In this chapter, I intend to elaborate this claim by exploring the interaction of performative and theoretical skills underlying what is arguably the most important epistemic activity in contemporary biology: the production of models. As observed in chapter 2, modeling is widely recognised to be a prominent way in which biologists obtain understanding of biological phenomena. However, the nature of the link between modeling and understanding is far from clear. Further, the relative significance of handling so-called ‘material’ models versus handling ‘abstract’ models has received much more attention among historians of science than among philosophers of science. In other words, the differences in the practices used and the results obtained by handling these two types of models are amply documented by historians, but are left widely unexplained by epistemologists of science.[148] What is precisely the difference between handling a material entity and handling an abstract entity? And in which (different) ways do these two types of activities contribute to the understanding of the phenomena to which these models refer?

Answering these questions proves especially difficult in the context of research on model organisms, which involves the combined use of a variety of models of different aspects of the biology of those organisms. I consider two types of models in particular which are widely used in Arabidopsis research: the two-dimensional, pictorial representations of Arabidopsis data displayed by TAIR (whose production is documented in the previous chapter); and the three-dimensional plant specimens cultivated and distributed by the Nottingham Arabidopsis Stock Centre. I devote my first section to reconstructing how NASC specimens (that is, the Arabidopsis specimens grown at the NASC) are

[148] See for instance the volume on three-dimensional models edited by de Chadarevian and Hopwood (2004), which, while providing wonderful examples of the contrast between ‘more material’ and ‘more abstract’ types of models, defines the difference between the two in little more than intuitive terms, without exploring its epistemological significance.

produced.[149] These practices can be compared to the activities required to obtain TAIR visualisations of Arabidopsis data, as described in chapter 5. As I argue in section 6.2, both TAIR images and NASC specimens should be seen as models of Arabidopsis wildtypes: the former because they provide symbolic, theoretically-informed representations of Arabidopsis (micro)biology; and the latter because they are modified to acquire and preserve specific phenotypic and/or genotypic features that are seen by researchers as representative of Arabidopsis plants. The claim that NASC specimens should be considered as models of Arabidopsis might seem confusing: on one hand, they are artefacts that are purposefully manipulated in order to represent that phenomenon; on the other hand, they remain samples of the very phenomenon that they are taken to represent. I argue that this ambivalence is precisely what makes these models so interesting and useful to understanding Arabidopsis biology, and thus is something that philosophers should try to account for in their discussion of modeling practices. However, the traditional categorisation of models as ‘abstract’ or ‘material’ does not seem to help in this respect. Many biological models, certainly including NASC specimens, present both abstract and material features: further, the meaning of this characterisation is unclear, as I shall show by listing three different ways in which the notion of ‘abstract model’ can be interpreted. A resolution of the debate on what it means for a model to be abstract would require a dissertation of its own. However, I do not think I have to provide such an analysis here: instead, I propose to shift the terms of the debate in a way that is more fruitful to the analysis of abstraction in the specific case of biological modeling.

In the context of modeling accounts, thinking of ‘abstract’ as an attribute of models black-boxes, rather than clarifies, the way in which abstraction is performed during model-building and to which epistemological advantage: it makes abstraction seem like an essential property of models, whose value is not dependent on the context and manner in which models are used. I thus suggest to view abstraction as an epistemic *activity* that needs to be skilfully performed in order to yield models of the phenomena that scientists wish to study and understand – an activity performed in different ways and for different purposes, depending on the local circumstances of modeling and the specific features of the models in question. This philosophical reading of abstraction leads me to qualify both TAIR databases and NASC specimens as models of Arabidopsis, whose differences in quality and expression are due to the difference in the processes through which they are built. These are two types of abstraction processes: the *material abstracting* required in the production of Arabidopsis specimens and the *conceptual abstracting* characterising the elaboration of visual models of Arabidopsis genomics and metabolism. The difference among these types of abstracting is determined by the epistemic goals, material circumstances (tools and experimental setting), background knowledge and, most importantly for my present purposes, the skills required by the scientists performing them. In section 6.3, I highlight how both material and conceptual abstracting require researchers to exercise theoretical as well as performative skills: however, while

[149] The material that follows has been gathered during a week-long visit to NASC in May 2005, in which I interviewed NASC team members Emma Wigmore and Lubomira Kacinova and had several in-depth interviews with NASC Director Sean May. Dr. May kindly granted me access to NASC archives as well as to its laboratory and glasshouses.

theoretical skills are predominant in the construction of models through conceptual abstracting, performative skills are especially important to the production of models through material abstracting.

These considerations lead me, in section 6.4, to propose a systematic analysis of how modeling practices help in coordinating knowledge so as to obtain a biological understanding of a phenomenon (thus clarifying and illustrating how my definition of understanding, as proposed in Chapter 2, actually works in biological practice). In particular, I shall highlight how the skilful handling of models enables researchers to select the (theoretical and embodied) knowledge that is most *relevant* to an understanding of the phenomenon that is being modelled – thus making the pursuit of understanding through modeling both skilful and efficient.

6.1 The Making of Plants: Producing Specimens at NASC

As I remarked in chapter 3, NASC and TAIR, despite their entirely different pursuits and results, share one important goal, that is, the pursuit of integration as proposed by the MASC committee supervising Arabidopsis research. Both centres have been created to produce tools that could be used by all Arabidopsis researchers, no matter their specific expertise and local interests: the use of similar tools would provide a material basis for exchanges within the Arabidopsis community and thus, eventually, for the integration of the various types of results produced by its participants. In chapter 5, I illustrated the impact of concerns about the usability of TAIR by Arabidopsis biologists on the making of the visualisations of data employed in the resource. TAIR images are meant to facilitate communication among Arabidopsis researchers and access to results from different research fields. Further, through the adoption of GO, TAIR provides a meta-framework where different sets of results can be compared and integrated with each other. NASC is equally committed to providing tools for integration of different fields. As I intend to show, NASC scientists also make reference to a meta-framework where different descriptions of the same plants can be brought to bear on each other; most importantly, they produce specimens that can be employed for research throughout the world and that are essential to the successful integration of results obtained across different laboratories.

The production of the ‘right’ plants for laboratory use is an extremely important task, as not all Arabidopsis specimens are fit for experimental purposes. Different branches of Arabidopsis research focus on specific ecotypes of Arabidopsis, whose features are particularly well suited to the experimental purposes at hand. As pointed out in Chapter 3, three ecotypes are particularly popular among Arabidopsis researchers working on molecular biology – the Columbia, Wassinlenska and Landsberg ecotypes (respectively abbreviated as Col, Ws and Lan). The main reason for this popularity is that researchers need some guarantee that the plants on which they work are as similar as possible to the ones used in other laboratories. Thus, plants used to obtain particularly influential results (such as the Col ecotype adopted by the AGI and the Lan ecotype used for Feldman mutations) are sought after by researchers wishing to assess, replicate and further those data. That plants share the same traits is essential to the reproducibility, evaluation and elaboration of experimental results across different laboratories: as we shall see, this is one of the reasons for viewing specimens of the most popular ecotypes as an important type of model of Arabidopsis. Further, integration of data about Arabidopsis can only occur if researchers have a reason to believe that they are working on similar plant specimens. This might look like a simple requirement, while actually it represents a great challenge for the Arabidopsis community, given the numerous variants of Arabidopsis to be found in the wild, as well as the wealth of mutants of the plant produced in the laboratory in recent years.

The production of Arabidopsis specimens with features appropriate to experimental purposes requires a complex sequence of activities, all of which are carried out at NASC. In what follows, I analyse these activities as pertaining to six different phases: the acquisition of Arabidopsis seeds from researchers (6.1.1); the preliminary classification of those seeds (6.1.2); the cultivation of specimens from those seeds (6.1.3); and the final classification, storage and distribution to the users (6.1.4).

6.1.1 Acquisition of seeds

Similarly to TAIR, NASC acquires its prime material – in this case, the seeds of plants in use in various laboratories – directly from its prospective users. Researchers who are working on Arabidopsis specimens taken from the wild or obtained through artificially induced mutation are encouraged to become 'donors': that is, they are encouraged to send a sample of seeds to NASC, so that the specific specimens can be replicated and stored in NASC archives and retrieved by anyone interested in working with plants with the same characteristics. Again, NASC works as a service to the community: a point of exchange for researchers working on Arabidopsis as well as an archive of the several hundred mutants of plants used for research in different areas and locations. Donors have an incentive to donate samples as a means to make themselves visible to the community. In fact, since the early days of Arabidopsis collections (such as the one compiled by Redei), the journals and newsletters dedicated to Arabidopsis research have promoted an ethos of sharing seeds and materials across the community, on the basis of the argument that cooperation would benefit all participants.[150]

The sample sent by donors is accompanied by a free-text description of the genetic and morphological characteristics of the plants from which the seed is taken. This is common practice among biologists, since there is as yet no standard terminology with which to describe the morphological features of a plant – the choice of terms and emphasis is bound to change depending on the use that researchers make of it, as well as their disciplinary expertise. For instance, consider the following definitions of the term 'bud':

1. *(botany)* A small protuberance on a stem or branch, sometimes enclosed in protective scales and containing an undeveloped shoot, leaf, or flower.
2. *(biology)* An asexual reproductive structure, as in yeast or a hydra, that consists of an outgrowth capable of developing into a new individual.

[150] More on this in Chapter 7, section 7.1.

3. *(medical)* The primordial structures from which a tooth is formed.[151]
The first two definitions are often used by donors to describe parts of Arabidopsis specimens, which generates confusion as to what precisely is meant by the description in each case.

6.1.2 Preliminary Classification: PATO

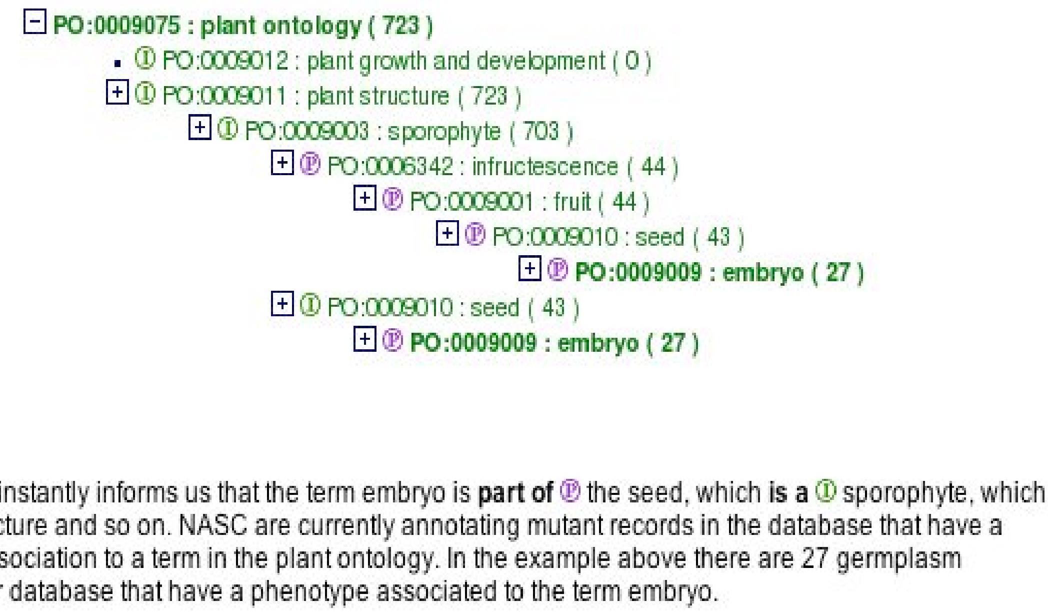
Given the diversity and ambiguity of descriptions of specimens provided by donors, the first thing that NASC researchers have to do when trying to categorise the incoming seeds is to re-phrase the description into a standardised description. This task is relatively easy in the case of the plant's genetic make-up, where terminology is more highly standardised and NASC can also make use of GO (and other classification tools provided by TAIR) in order to adapt the description to a common standard that is universally accessible and adds coherence to the NASC storing system. Standardising descriptions of the plants' morphology proves much more problematic, given the great diversity and inaccuracy characterising descriptions of plants at that level (as in the example above).

To solve the problem, NASC is developing its own system of classification under the name of Phenotype, Attribute and Trait Ontology [PATO]. This system is structurally similar to the GO system, insofar as it also exploits the child-parent network characteristic of Java programming (like GO, PATO is part of the umbrella project of Open Biological Ontologies, as mentioned in chapter 5). The difference between PATO and GO lies in the type of concepts as well as the type of relations used in the network, which of course depends on the different goals set by each classification system. Concepts in PATO are mainly 'attributes' describing morphological features of the organism in question or 'components' of the organism. 'Traits' constitute the values and/or properties characterising these attributes. For an example of how this preliminary classification of plants works, consider the case of a set of seeds arriving at NASC together with the following description: 'Col ecotype, glabrous leaves'. PATO allows NASC researchers to assign a value to each component of this description: the entity being described ('leaf'), the attribute being considered ('pilosity') and the value given to such attribute in the case of this specific Col specimens ('glabrous'). By assigning a precise definition and serial number to all attributes and components, PATO successfully standardises the descriptions available for each Arabidopsis specimen, while also resolving the ambiguity intrinsic to many of the relevant terms.

Given the anatomical focus of PATO, concepts are classified through a top-down, mereological approach starting from the whole organism ('plant') and branching out to its components ('fruit', 'seed'). Although the same relations are used as in the case of GO ('is_a', 'part_of' and 'develops_from'), the privileged relation here is clearly the mereological relation 'part_of', as the descriptions of plants start from the morphological level ('tissue') and connects it ('part of') to relevant components ('cortex') (figure 6.1).

[151] NASC website accessed 20 April 2006.

Figure 6.1 - Example of how PATO works, as provided by the NASC website (accessed 20/04/2006). Here the green symbol I indicates an identity relation ('is_a'), while the purple symbol P indicates a mereological relation ('part_of').



6.1.3 Cultivation of specimens

Once the seeds are classified according to PATO standards, NASC researchers grow plants out of them so as to gather a sufficient quantity of seeds for storage and distribution purposes. Like most plants, Arabidopsis can self-fertilise as well as crossbreed with other plants. A single plant can produce over 10.000 seeds and each of these seeds will have precisely the same characteristics of the 'mother' plant, unless the plant is fertilised by seeds from another variant (in which case recombination occurs, generating an offspring sharing the traits of both parents). Provided that it is cultivated in strict isolation, growing Arabidopsis in its haploid mode is the most efficient method for multiplying seeds of the same variant. It also allows to check that the description provided by the donor is correct and to determine the best conditions for the growth of the specimen in question (protocols indicating how to grow each specimen are given to prospective users together with the relevant seeds, so as to make sure that plants will grow in the same way as in the NASC glasshouse and will therefore have precisely the same features and dimensions).

Cultivation is the most labour-intensive phase of specimen production. It includes the following processes:

1. *Germination*: the seed are sown in a special container facilitating their germination (figure 6.2);
2. *Growth*: the plants (usually around 25 per type, but sometimes as few as 1 per type, depending on the popularity of the variant) are grown in a glasshouse

equipped with humidity and temperature control and an isolation system preventing cross-contamination (figure 6.3);

3. *Drying*: the plants are transferred to a second glasshouse containing ventilators, in order to dry them as required for the extraction of seed;
4. *Harvesting*: dried plants are squeezed and pulverised in order to gather the seed (figure 6.4); seeds are then separated from other plant remains and eventual dirt or dust from the glasshouse via a process of sterilisation.

Figure 6.2 – NASC technician sowing a batch of seeds. The number and distribution of seeds is gauged by hand.

Figure 6.3 - NASC glasshouse for growing specimens (the visible ecotypes are of the Lansberg variety, the most popular ecotype for laboratory use together with Col and Ws).

Figure 6.4 – Harvesting of seed (before sterilisation)

As noticeable from the photograph depicting the NASC glasshouse (figure 6.3), these processes are carried out in controlled environments via tools standardised to that effect.

The growth of plants has been automated as much as possible, so as to avoid variations in handling deriving from human intervention. On the other hand, this extended automation still requires a great amount of manual curation by NASC personnel. No precise techniques or tools are used to count or weigh seed, which means that technicians evaluate the right amounts to be sown or harvested by reference to their previous experience rather than to guidelines or measurements (figures 6.2 and 6.4). It also happens that some plants show signs of ill health, due to mishandling, parasites or unexpected mutations (generated by unwanted cross-breeding) making plants unable to cope with the living conditions enforced in the glasshouse environment. These cases are noted and the plants in question are transported to a separate room, where each plant is taken care of on an individual basis (figure 6.5).[152]

Figure 6.5. Refrigerated room where plants in need of special attention are stored and cared for. Each of these plants is illuminated, cut and watered according to its present needs (as signalled on the yellow reminder stuck on them).

[152] The discarding of plants judged to be ill or anyhow inadequate to the standards set by NASC is a good example for the need, by NASC modellers and Arabidopsis researchers alike, to have 'right tools for the job', as put by Clarke and Fujimura (1992).

In sum, NASC researchers need to evaluate the growth rate and conditions of plants, intervening in the harvest whenever needed, multiple times per day. They acquire an extensive range of performative skills and, more generally, embodied knowledge about the plants that they are handling: actually, and despite the attempts to provide as much information as possible on how to handle the plants to the prospective user, it is very difficult for NASC to convey to others the amount of embodied knowledge accumulated by its technicians and researchers.

6.1.4 Final Classification, Storage and Distribution

Both the genetic make-up and the morphological features of the plants whose seeds are now secured in high quantity are double-checked, in order to ensure that the information used to label the seed samples are accurate. Seeds are cleaned and sterilised, after which a couple of thousands of them are frozen into the NASC 'seed archive' (so that a copy of any classified type of seed is always retrievable, no matter how many seeds are distributed to Arabidopsis researchers around the world). The rest is stored in appropriately labelled sample packages, ready for distribution.

Seeds are put into sterile containers and shipped to the prospective users upon request, together with guidelines on how to grow and handle them. Remarkably, TAIR and NASC are directly related in this phase, since the TAIR website allows users to order from NASC precisely the specimen on which the data visualised through the resource have been obtained in the first place. In addition to NASC's own website, TAIR has therefore become a very important portal for ordering NASC specimens – a practical demonstration of how close in purpose and organisation these two apparently unrelated enterprises really are.

6.2 Modeling a Model Organism

The extensive list of interventions required to produce NASC specimens highlights the fact that the transfer of Arabidopsis specimens from the wild to the laboratory involves more than a change of context (though the latter certainly has great impact on the representational value of the specimens). Obtaining specimens that are adequate for research purposes requires apposite structures, standardised tools and guidelines, plus extensive experience in handling the plants as well as experimenting on them: in other words, it is a matter of *skilful production*, rather than simple displacement. By the time that a seed sample is labelled and sent off to a user, researchers have achieved a high degree of control over the plants that will grow from those seeds.[153] The PATO description summarises and defines which traits the plants are supposed to grow. Those

[153] For instance, the leaf shape characterising the Columbia ecotype is replicated across Col specimens by gaining control over their genetic make-up and the environmental conditions under which they are grown (the possibility of control is greatly enhanced by Arabidopsis being a self-fertilising organism).

traits, which are selected depending on the experimental needs of the scientists using the plants, are then reproduced under the material constraints provided by the laboratory equipment and the resistance to modification by the organisms. The resulting plants are hybrids of 'wild' traits and traits that are well controlled by experimenters and thus stabilised and reproducible.

To shed light on the significance of this ensemble of practices, as well as to connect them with the practices characterising the making of TAIR that I explored in the previous chapter, I shall now proceed to compare the epistemic activities involved in the production of NASC specimens with the epistemic activities used to produce TAIR visualisation tools. This comparison turns out to illuminate a variety of issues concerning the modeling of model organisms as well as its significance towards the understanding of such organisms' biology. Indeed, I wish to suggest that both series of activities constitute (two different types of) modeling practices. These activities are successful, insofar as they result in a wide range of models of specific aspects of Arabidopsis biology.

6.2.1 TAIR Images and NASC Specimens: Two Types of Models

The visualisations of data to be found in TAIR constitute exemplary cases of two-dimensional, visual models of Arabidopsis. They clearly encompass a lot of information about specific aspects of Arabidopsis biology – mainly concerning its molecular structure and functions. In this sense, they are *representative of* a series of phenomena. At the same time, they are structured by reference to a theoretical framework (the GO) and are therefore *representative for* that framework.[154] Further, this double role does not exhaust their epistemological status. By virtue of the specific skills and tools employed in their construction, neither of which are necessarily related to the theories and phenomena that the models are taken to represent, these images acquire an independent, autonomous status from such theories and phenomena: they start a 'life of their own' (Morgan 2005, 318). In this sense, each of this images can be viewed as a mediator between theory and the world, according to the interpretation provided by Morgan, Morrison and their associates: TAIR images constitute models of specific aspects of Arabidopsis biology.

Interestingly, the multiplicity of these images and the different ways in which they portray data about the same phenomenon (as we saw, there can be different visualisations emphasising the processes, components or functions associated to each Arabidopsis trait, such as metabolism or gene products) do justice to the need for model pluralism that I advocated in section 2.2.2. As we saw there, the creation and use of a plurality of models of the same phenomenon is not only useful, but actually necessary to its study: each type of model captures a different aspect of it, thus contributing to an overall explanation of that phenomenon. All of these models have to conform to the level of generality imposed by the adoption of GO as a theoretical framework: data visualised in those images are first associated with specific GO concepts, then interpreted in the light of the whole GO network to which those concepts belong, and only then can they be employed in a diagram depicting Arabidopsis microbiology. Thanks to this extensive theoretical

[154] See Chapter 2, section 2.2.2.

manipulation, these images have high *explanatory power*: they do provide a vision about how specific data can be interpreted in order to account for the specific feature or behaviour of the organism in question. Such a vision allows researchers to envisage the potential empirical significance of a theory, without however providing means to test whether such potential explanation actually applies to phenomena.[155] As we saw in the previous chapter, a correct reading of such a vision is only possible thanks to adequate theoretical skills. Further, biologists willing to exploit the explanatory power of these models need to possess relevant background knowledge, both theoretical and embodied in kind: they need to know about GO, for instance, in order to learn the theoretical skills allowing them to use the models efficiently for their own research purposes; and they need to know protocols and guidelines for using specific instruments and research techniques, so as to learn the performative skills enabling the application of these models to the investigation of the actual phenomena under scrutiny (in this case, the plants). In fact, as clearly indicated by Morgan and Morrison, researchers can learn something from a model only when they are able to manipulate it and modify it to suit their research objectives. In order to handle TAIR images, and thus profit from their explanatory power, Arabidopsis biologists need to know what these models are supposed to represent as well as being able to think about those representational assumptions and modify them to suit their own interests. For instance, a model of Arabidopsis metabolism (figure 5.12) can be usefully manipulated when (1) users of the model refer to labels provided by TAIR (indicating what different components of the model stand for, as in 'red triangle = protein') and their own background knowledge on plant metabolism; (2) users possess theoretical and performative skills allowing them to play around with the model, handle and modify it according to their critical judgement of such labels and background knowledge.

A notable feature of this description of TAIR models is that their explanatory power springs from background assumptions, labels and concepts used to interpret them, rather than through embodied knowledge of the phenomena to which these models refer. Such embodied knowledge is, as we have seen in the previous chapter, indispensable to the construction of these models and, arguably, to their use: yet, it is possible to manipulate the above model of Arabidopsis metabolism and learn something from it without having any embodied knowledge (least of all performative skills) of how Arabidopsis plants actually develop under varying environmental conditions. In fact, these digital models provide little or no feel for the *empirical adequacy* of the data that they contain: in order to evaluate what these images teach us about Arabidopsis biology we need to complement them with the embodied knowledge acquired through experimental intervention in actual plants. One way to do this is to complement the use of TAIR images with other types of models, which would be better suited to account for the empirical import of the theoretical suggestions contained in TAIR digital models. As indicated by Richard Levins already some decades ago, the combined use of models of

[155] In some cases of theoretical modeling, explanatory power does not even involve potential empirical significance. For instance, a mathematically complex model of an evolving population might provide some interesting theoretical insights (such as the rate of potential population growth under specified conditions): yet researcher using that model might not know how to apply it to any real biological population, as the specified conditions favour mathematical elaboration but are actually inexistent within the natural world.

different types (such as, in this case, models with high explanatory power and models that are empirically accurate) constitutes a convenient trade-off towards productive research on biological phenomena. I now wish to propose NASC specimens as a type of model that might successfully complement TAIR images towards achieving a specific goal: the combined use of these two types of models constitutes a good strategy towards acquiring an integrated understanding of Arabidopsis biology.

I shall elaborate this claim in section 6.3. Before doing that, however, I need to examine a fundamental premise for this argument: the claim that specimens produced in NASC function as models of Arabidopsis. This claim turns out to be trickier and more controversial than in the case of TAIR images, and I shall therefore devote the rest of this section to it. The epistemic status of the organisms employed as model organisms in contemporary biology has been widely discussed by several historians and philosophers. It has been argued that such organisms are no models at all, but rather that they are simply the *materials* on which experimental interventions are carried out. Others have claimed those materials to be *samples* of the phenomena under scrutiny and that that does not amount to recognising their role as models since, again, 'real' science lies in the 'abstract' modeling of these objects.[156] In contrast with these views, I do not think that there should be doubts about the representational value of NASC specimens as models of Arabidopsis wildtypes. This is because the standardisation and controlled distribution of Arabidopsis seeds has a direct impact on the content of the knowledge produced by using those plants. As we have seen when surveying the making of NASC specimens, the plants are cultivated to preserve specific traits, both at the genotypic and at the morphological levels: as indicated in section 6.1.3, plants that turn out not to conform to these standards are taken out of NASC glasshouses (and thus of their distribution circuit) and they are either destroyed or grown separately from the rest of the harvest. NASC researchers acquire a great deal of control over the features of the plants that they want to reproduce and distribute to Arabidopsis researchers. This means that the results acquired through the study of these plants are tightly linked to their specificity as products of the NASC selection process. This makes them at least partially autonomous from the Arabidopsis wildtypes festering in the European countryside. NASC specimens have been modified to fit some specific theoretical expectation and experimental requirement, with the result of producing entities that are at the same time related and partially independent of both the phenomena and the theoretical issues which they are used to represent and explore: like TAIR images, NASC specimens become mediators between theory (including the researcher's expectations and interests) and the phenomena that they embody, that is, Arabidopsis plants.[157]

---

[156] Rachel Ankeny implicitly supported such a view in her (2000; 2001) papers on the modeling of C. elegans, where she proposed that what counts as a model of C. elegans in neurological research is what she calls the 'wiring diagram' representing the neural connections characterising the nematode. As she noted, this meant reducing the model organism to an idealised, abstract entity. Ankeny recently responded to my critique of this aspect of her work by retracting this idea and providing a much richer analysis in her (2005; 2006), which highlights both the material and the abstract features of model organisms as models.

[157] One powerful objection could be levelled against the argument that model organisms constitute models for biological phenomena in certain research contexts. This is the idea that what scientists use as models for biological phenomena are various components of the organism, rather than the organism as a whole. Indeed, most Arabidopsis researchers do not need to examine the whole specimen of the plants they

In fact, the representational value of NASC specimens as models often transcends the class of weeds to which the original Arabidopsis plants belong. Depending on the goals and resources of researchers experimenting on NASC specimens, these models are often assumed to hold for all plants, irrespective of the highly controlled conditions in which these plants are produced as well as of the huge differences characterising the tens of thousands existing species of plants. This is a very important assumption, as it determines the representational value of NASC specimens as models: the wide applicability assigned to the results gathered from NASC specimens transforms these plants into *models of any plant*. Of course, not all results gathered on Arabidopsis specimens are given the same representational value. There are cases where research on NASC plants yields results that are taken to hold only for Arabidopsis ecotypes, if only because the applicability of those results to other organisms still has to be tested. For instance, plant immunologists are yet investigating the significance of their studies of Arabidopsis/pathogen interactions for immunological responses in other plants. In other cases, the representational value of data acquired through these specimens soars, as for example in the case of recent findings on mitochondrial DNA[158]: the validity of these results is taken to extend to all plants and animals, including human beings, since plants and animals are assumed to display the same molecular mechanisms. It thus becomes apparent that what NASC specimens are *models of* changes depending on the research context in which results acquired through experiments on these plants are used.

I should note here that the special representational value bestowed on Arabidopsis in the last few years does not in any way imply a *belief*, on the side of the scientists making that assumption, that one plant can indeed be taken as a representative of the whole floral kingdom. The current epistemological status of Arabidopsis specimens is, rather, dictated

---

investigate, but only a small sample of tissue or a specific cell culture. It might be argued that, in those cases, it is only those specific cells, and not the whole plant, that should be regarded as models of, for instance, a given biological mechanism or pathway. I do not see this claim as incompatible with my view of model organisms as models. I agree that, within certain settings, components of Arabidopsis specimens can be usefully regarded as material models of specific biological phenomena (they are what Arabidopsis biologists call 'model systems'; interview with Sue Rhee, 18 August 2004). However, there are very good reasons for regarding whole plants as material models, too, especially when the phenomenon that is modelled is a complex one, involving various plant components and different levels of biological organisation. When 'producing' Arabidopsis specimens, NASC technicians have three ways of modifying their traits to fit the expectations of Arabidopsis researchers: the first is to modify their genome via interference through *Agrobacterium* or knock-out experiments; the second is to breed them as required to obtain specimens with the desired features; and the third is to control their development by controlling the environmental conditions in which they grow. In fact, there is no known way to produce components of the plant independently of the rest of the organism: what NASC technicians aim at reproducing are the morphology and genotype of whole plants, not their singular components, *precisely because those components are working together in ways that are not yet understood*. Taking a whole plant as a model for a given phenomenon (for instance, a transcription mechanism) means retaining awareness of how several parts of that model at different levels of organisation (e.g. enzymes, RNA, proteins, cells) may interact in order to produce that phenomenon. In biological research, it is increasingly important to consider the organism as a whole in order to understand any one phenomenon characterising its biology (good examples of this being the epigenetic regulation of gene silencing and plant responses to environmental stress, such as responses caused by pathogens or changes in temperature).

[158] See review by Millar et al (2004).

by the practical necessity of focusing research efforts on one organism. In fact, Arabidopsis researchers – not to mention biologists working on other model organisms – do not hesitate in indicating the many phenomena that Arabidopsis cannot be representative for. These include, for instance, RNA interference (injecting RNA into embryo to see what happens), which can be done on petunias but not on Arabidopsis[159]; homologous recombination (which is better done on yeast); and several studies of plant pathogens, which would require an organism with a much longer life span than Arabidopsis.

This recognition does not take anything away from my argument that these plants constitute models of (often very broad) classes of organisms. In fact, I argue that the use of highly standardised Arabidopsis specimens, rather than specimens with different characteristics or of a different species, has great bearing on the outcomes of research on plant biology: the handling of NASC specimens, rather than natural variants, has a strong influence on the procedures adopted to work on the plants, as well as on the results obtained from such research and their interpretation. One way to illustrate this point is to look at how mistakes in the procedures used to produce NASC specimens can affect the use of specimens as models and, as a result, the biological knowledge derived from them.

Consider the case of C24, an ecotype that enjoyed a brief but high popularity among Arabidopsis researchers at the end of the 1980s. Sean May, the director of NASC, defines this case as an 'example of what could happen when you don't have a stock centre' supervising the production and distribution of seed.[160] C24 acquired a good reputation as a very easily transformable ecotype in the 1980s and was thus privileged over Lan when it came to experiments involving mutagenesis. Importantly, it was classified as a mutant of Col (hence the 'C') and results obtained in reference to it were interpreted accordingly. Upon further research, however, it became clear not only that all main ecotypes were easily transformable (thus eliminating the main reason for the popularity of C24), but that the putative identity of C24 as a mutant of Col was altogether a mistake. A study of its flowering time alleles reveals that C24 and Col are polymorphic, and thus do not share a common heritage (Sanda and Amasino 1995, 2). C24 turned out to be a descendent of the wildtype Coimbra, which has been wrongly classified because someone had written down its name in an illegible handwriting ("Coimbra' thus turning into 'Columbia'). This mistake in classification did have consequences for research. Researchers were insisting on the use of C24 because of its (non-existing) genotypic closeness to Col, despite strong evidence that C24 had strong disadvantages to its use as a model for other plants. The most serious among those was its being carrier of two mutations that, when separated by interbreeding, produced late flowering – in other words, that transformed weed into cabbage! Following the uncovering of C24's real identity, several experiments carried out and interpreted on the assumption that C24 was similar to Col had to be re-examined and, in some cases, discarded. Arguably, this seemingly blatant misunderstanding survived well into the 1990s precisely because, when C24 was initially adopted, the Arabidopsis stock centres did not yet have the function of checking upon the lines used by

[159] The reasons for this are unknown to biologists, one of whom remarked to me that 'it's meaningless to ask why, actually, it's like asking why it is so easy to transform Arabidopsis'.
[160] Interview, 16 May 2005.

researchers. In the absence of an institution devoted to such double-checks, errors of this kind are all too likely to be perpetuated in biological literature (which is often focused on the production of new data rather than on the verification of old ones).

Other types of problems can be caused by mistakes in devising protocols and tools for the cultivation of plants in the laboratory. An instance of this is the slowness in improving strategies and tools that ensure that plants of different strands can be grown in the same room without risks of cross-contamination. Back in the 1980s, specimens of a single strand were grown on gravel in 25 exemplars per strand, and plants of different ecotypes would be separated from each other by plastic walls. This solution turned out to be both space-consuming and risky because contamination could still happen through the air. The next step was then to cultivate only one plant per strand (which would anyway produce thousands of seeds by itself), grow them in separate pots and isolate them from the environment by attaching a stiff tube of plastic to the base of the pot, which would reach up to thirty centimetres over the maximal height of the plant. This second solution was also a failure: while efficient in shielding contamination by circulation of seeds in the atmosphere, the tube did not allow plants to grow to their full potential (especially Columbia specimens, which are much shorter than Landsberg erecta and whose leaves are round and big, needing more horizontal extension). Sean May came up with a better technology by thinking about ways of packaging flowers in supermarkets and florist shops. NASC now uses ‘zwapak flower sleeves’, a type of plastic bag exported from a Dutch flower shop, which is efficient in shielding contamination, flexible enough to let the plants breathe and grow to their full potential, and can even be bent and crushed at will for a better germination and easier harvesting of seeds (figure 6.6). Even those, however, have some problems: when I visited the NASC glasshouse, some of the packs had been bent by the ventilation system and actually crushed the plant – not an encouraging sight, when considering that each pot constitutes the unique sample of a whole mutant line (figure 6.7).

Figure 6.6 – Pots isolated by zwapak flower sleeves. Note that the specimens in the pots are already dried out and ready for the harvesting of the seeds.

Figure 6.7 – Flower sleeves crushed by the ventilation system. The risk is then for seeds of the plants contained in the pots to disperse and mix with seeds from other plants.

6.2.2 The Ambivalence of Specimens as Models

From these examples, it becomes evident how the specific traits of NASC specimens (as grown via appropriate handling and cultivation procedures) are influential to the production of biological knowledge on plants, since those exact traits are assumed, by homology, to represent traits in several other (often widely different) organisms.
The latter example also illustrates the extent to which the handling and growth environment of specimens influences their features. By increasing their expertise on the maintenance of plants that best fit their purposes, NASC researchers acquire a great deal of control over the plants' morphology and growth. Arabidopsis is not alone in this respect: most widely used model organisms present characteristics that are rarely, if ever, found outside of the lab. In his illustrious study of Drosophila melanogaster, which also highlights the high degree of control acquired by geneticists over the fruit flies, Kohler uses this point to argue that model organisms are, in fact, *artefacts* of biological research. He claims that, as in the case of Arabidopsis specimens, laboratory organisms are reproduced and modified under such controlled and purpose-oriented conditions that they end up bearing little resemblance to their relatives in the wild: their features have been largely reshaped by scientists according to their research interests - a fact which, in Kohler's eyes, signals that laboratory organisms should be considered as artificial products of human interventions rather than samples of nature (Kohler 1994).

Kohler's point, if taken seriously, has great consequences for the representational value of model organisms as models. The idea of model organisms as samples of nature is often used to justify the representational value of those organisms with respect to their relatives in the wild: the idea that experimental manipulation might compromise such representational value is not contemplated, as it would nullify the scope (and thus, the significance) of the results of research carried out on them. Given this context, Kohler's suggestion is valuable and provocative, because it forces biologists to put into question the extent to which model organisms actually represent organisms growing outside of the laboratory. This is a highly contested issue in model organism research, as I briefly mentioned above and in chapter 3. I already highlighted the dubious epistemological status of model organisms research, insofar as the obvious advantages of being able to accumulate knowledge on just a few organisms might be outbalanced by the lack of justifications for extending the applicability of such knowledge to the millions of species that are only related to those models through common descent. It is important to keep in mind that the very idea of producing NASC specimens for distribution across the globe has deep roots in microbiology and especially in molecular biology. Ecologists and even some developmental biologists are sceptical of the possibility to study the huge diversity characterising organisms on the basis of a few exemplars. This approach is favoured by microbiologists, who argue that the 'basic' mechanisms of genetic transmission, biochemical processing (such as involved in cellular communication and metabolism) and cell biology are common to most organisms. Even when admitting the validity of this argument (which has itself been contested), it displays a strong bias in favour of molecular biology in the context of model organisms research. Having several specimens of the same ecotype is useful to investigate mechanisms and micro-structures common to all plants; it is much less useful in investigating ecological features, such as comparing the development of different ecotypes under the same environmental conditions. Thus, the special representational value assigned to NASC specimens over and above natural variants has been extensively criticised by ecologists and other biologists interested in natural variation among Arabidopsis species. These scientists argue that reliance on standardised and controlled specimens of plants prevents researchers from taking into account the influence of the environment, cross-fertilisation and specific ecological niches on the structure and development of Arabidopsis traits (Alonso-Blanco and Koornneef, 2000). By pointing to the number of Arabidopsis features that are not representative for plants in general, opposers of the use of NASC specimens agree with me that these plants are used as models of plant biology: in fact, they criticise this use as unjustified and potentially misleading, insofar as the representative value attributed to these models is often too broad ('all plants' rather than 'this specific family of Arabidopsis ecotypes').

In this thesis, I do not wish to take sides in this debate. There is a simple reason for this choice: I do not think that research on NASC specimens should be condemned, just as I do not think that it is useful for any type of investigation in plant biology. Of course, for the purposes of comparative research (especially involving ecological aspects), the use of standardised specimens is bound to be misleading rather than helpful, since the manipulation of specimens aims at reducing the natural diversity among them. What I wish to emphasise here is that there is a strong sense in which NASC specimens are very useful models in order to enhance the biological understanding of plants: that is, there is an undeniable sense in which they are actual samples of the phenomena that they stand for, rather than artefacts produced to represent those same phenomena (as Kohler would have it). NASC specimens are, after all, actual organisms: entities that we could not hope to create from scratch in a laboratory (despite many attempts in this direction, e.g. robotics), precisely because we do not know, and understand but a minimal part of, their functioning. Organisms, even when highly controlled as in the case of NASC models, retain a mysterious quality that is crucial to their role in the laboratory: they are actual samples of the very phenomena that they stand for. As I emphasised in my description of modeling activities carried out at NASC, popular model organisms used in actual laboratories, such as Arabidopsis, go through an extensive process of preparation. In this sense, the organism selected to be a 'model organism' is certainly not a random sample taken out of its natural environment. The transition from natural to laboratory environment is accompanied by a series of modifications of the organism itself. The process of preparation of an organism for experimental use consists precisely in a selection of traits that researchers wish to focus on (and thus, to stabilise and control). On the other hand, these traits are parts of a whole material organism – an organism that is partly 'built' to fit research purposes, but that preserves components that are untamed and sometimes even unknown to researchers. It is important to keep in mind how model organisms are not only typical of a phenomenon, but also 'are', to an extent, the phenomenon itself.

This intuition is well illustrated in the following passage from Keller's discussion of the role of model organisms in biological research: 'unlike mechanical and mathematical models (and this may be the crucial point), model organisms are exemplars or natural models - not artifactually constructed but selected from nature's very own workshop' (Keller 2002, 51). While I disagree on Keller's outright rejection of Kohler's point, I welcome her attempt to re-focus on the features of model organisms that make them into a special, and indeed precious, type of model in experimental biology. How to put together these two, apparently opposing, insights about the nature of model organisms as models? My view is that we should recognise the ambiguous status of model organisms – which are taken to be at once samples of nature and human artefacts, abstract and material entities, known and unknown phenomena – as precisely the feature that makes them such interesting and important objects in biological research. Model organisms are, indeed, models of the 'untamed' class of organisms of which they are a sample (and can even be taken as representatives for a much wider class of organisms, as in the case of Arabidopsis plants becoming representative for all plants in certain research contexts) – and yet, they are highly domesticated samples, whose handling and traits are so familiar to the researchers employing them as to become 'tame', that is, subject to control by those researchers. In fact, work on model organisms implies a *reciprocal domestication* of scientist and phenomenon: the scientist learns to handle and know the phenomenon by modifying it, so that the phenomenon itself adapts to such handling by changing some of its distinctive features.[161]

Oddly, perhaps, the best illustration for this process of reciprocal domestication does not come from philosophy but from literature.[162] In his famous story about the Little Prince, Antoine Saint-Exupery devotes a chapter to describing the blossoming friendship between the prince and a fox. The fox asks the prince to be tamed. Upon the prince's question 'what does that mean.. tame?', the fox replies:

> "It is an act too often neglected," said the fox. "It means to establish ties. [..] To me, you are still nothing more than a little boy who is just like a hundred thousand other little boys. And I have no need of you. And you, on your part, have no need of me. To you I am nothing more than a fox like a hundred thousand other foxes. But if you tame me, then we shall need each other. To me, you will be unique in all the world. To you, I shall be unique in all the world. . ." (Saint-Exupery, ch.12)

[161] Harré (2003) also uses the notion of domestication to describe model-researcher interactions. However, I take distance from his account, within which the term domestication is used to define a specific type of simplification (ibid., 28), rather than the more complex processes of standardisation and abstraction (leading to actual modifications in the nature and properties of the object at hand) that I consider here.

[162] The intuitions underlying the notion of reciprocal domestication are the same underlying the biological mechanism of co-evolution. Donna Haraway attempts a philosophical discussion of co-evolution when considering the ontologies of companion species (such as dogs and cats), which she defines as 'figures of a relational ontology, in which histories matter; i.e. are material, meaningful, processual, emergent and constitutive' (2003, 69). She does not, however, explicitly analyse how such co-evolution of humans and other life forms happens, nor its consequences (particularly in a context such as model organism research).

The fox goes on to argue that taming occurs by iterative encounters, during which the two parties get to know each other and acquire habits (‘rites’) through which to interact with each other. The acquisition of these habits changes the way in which both parties behave: just like, in the constrained environment and standard practices characterising interactions in the laboratory, the repeated encounters between scientist and phenomenon results in modifications of the features and behaviour of both. It is precisely these modifications that allow for reciprocal knowledge: the plant learns to respond to certain instruments and tools (such as the Agrobacterium inducing mutations), while researchers learn to observe and manipulate plants so as to obtain the expected results.[163] Organisms functioning as models are both found and produced; and both their selection and their production require the research community that employs them to acquire an appropriate ethos, structures and resources.[164] Plants thus become models of a specific phenomenon and/or issue, while biologists acquire the embodied knowledge required to handle the model. This results in the balanced mixture of the natural and the artificial that characterises model organisms as models. As a sample, model organisms such as NASC specimens are highly domesticated so as to become representative of a whole class of organisms (including other variants of Arabidopsis, as well as many other plant species); yet, they are only partially an artefact of human intervention.

Reciprocal domestication accounts for the combination of *natural*, *induced* and *projected* features that I listed as characterising model organisms in chapter 3. Further, the procedure of reciprocal domestication echoes Latour’s ideas about the representational status of laboratory results: findings obtained through the study of phenomena brought into a laboratory do not tell anything about nature outside of the laboratory, since natural phenomena can only be brought into an experimental setting through substantial intervention on (and thus modification of) their original features (Latour 1988). On the other hand, my framework does not deny that laboratory findings might provide us with an understanding of ‘untamed’ nature: in fact, I do think that studies of model organisms can yield results that are applicable to plants living in the wild. What I am pointing at is the necessity to acknowledge the ‘artificial’ nature of the models employed in the laboratory (including model organisms) in order to evaluate their representational value with respect to natural phenomena. The representational value of these model organisms lies precisely in the ambivalence characterising their epistemological status.[165] NASC specimens are, on one hand, produced and marketed by a group of researchers in order to represent some specific features of Arabidopsis biology - such as for instance the genotypic sequence embodied specifically by the Columbia ecotype (a sequence that is not, or anyhow rarely, shared by other Arabidopsis ecotypes). They are artefacts, in the

[163] Of course my use of the verb ‘learning’ here is partly metaphorical, as learning for plants is in many ways a very different process than for humans.

[164] The importance of community ethos has been argued also in the case of Drosophila: 'standard flies were not just the means of experimental production but also the bearers of a distinctive moral economy and a distinctive way of experimental life' (Kohler 1994, 168). Without some communality of ethos, funding sources and communication tools, the skills and networks needed to produce and distribute Arabidopsis specimens would not exist: I come back to this point in chapter 7. See also Leonelli (2006).

[165] I here interpret the term ambivalence etymologically: the term comes from the Latin *valent*, meaning valuable, and *ambi*, meaning ‘in two ways at once’.

sense of being selected and bred so as to classify as a 'normal' specimen of a specific class of organisms. At the same time, they are actual samples of those organisms, thus retaining much of the mystery and wilderness of their undomesticated relatives – that is, the several types of weed infesting European and North-American countryside and cities. This latter feature allows researchers to use these organisms for exploring, and eventually discovering, aspects of Arabidopsis biology that are yet unknown.[166]

6.2.3 A Second Look at 'Abstract' Models

Given the ambivalence of NASC specimens as models, what can we say about their features, especially when comparing them to the features characterising TAIR images? A first temptation would be to declare, with Kohler, that the features of these models are purely material: they are three-dimensional entities whose epistemological value is determined by the possibility, on the side of researchers, to physically intervene in them. This notion of 'material model' has become a popular topic in the philosophy of biology, thanks to the work by Griesemer (1990), de Chardaverian (2003) and Rheinberger (1997): yet, disagreement remains on which features of these models, if any, are abstract. The problem springs from the fact that, in order to be models at all, these material models have to be representative of a broader class of organisms and representative for a set of issues or processes: these attributions of representativeness are in themselves abstract qualities, since they are not material features of the models, but they are endowed on them by the researchers. While developing my views on the ambivalent – and, hence, valuable – status of model organisms as models in biology, I have come to regard the debate on whether certain types of models possess material and/or abstract features as a peculiarly unfruitful way to analyse the epistemological advantages characterising the handling of models. The notions of abstraction and materiality are treated as essential and often incompatible attributes of specific models: the recognition of a model as 'abstract', for instance, is supposed to tell us a lot about its characteristics and the way in which it is used in order to obtain scientific understanding. Yet, there is no consensus within the relevant literature on what it means for a model to be abstract: in fact, I can list at least three (arguably contrasting) intuitions underlying the use of the term 'abstract' as an attribute of models.[167]

One such perspective concerns the extent to which models are embodied in objects that can be perceived via our senses. In this first sense, the adjective 'abstract' figures as the opposite of 'concrete' and 'material': something is abstract when it is not tangible and/or visible, as in the case of a mental construct or an imaginary object for instance. This is an absolute notion: there are no 'degrees' of abstraction, as an entity is either tangible or not. According to this interpretation, neither TAIR images nor NASC specimens are abstract,

[166] Rheinberger's analysis of experimental systems resonates with this second epistemological role of Arabidopsis specimens, when he notes that '[experimental systems] are not simply experimental devices that generate answers; experimental systems are vehicles for materialising questions' (Rheinberger 1997, 28).

[167] My list is not supposed to be exhaustive, but rather to give an idea of the confusion underlying the use of the term 'abstract' in the discussion of modeling practices.

since they are both embodied into an image on a computer screen (the former) and an actual organism (the latter).

A second interpretation concerns the amount of information conveyed by a model with regard to its empirical import or applicability, that is, the way in which it relates to the phenomena that it is taken to represent. Thus, a model is the more abstract, the more it is devoid of physical meaning.[168] Note that this notion of abstraction admits of degrees, which can be measured by the amount of additional information needed for a model to be applicable to the analysis of a particular phenomenon. This second intuition conflicts with the previous one both in its motivation and in its implications. There is a clear difference in the amount of additional information needed to relate TAIR models to the appropriate component of Arabidopsis plants, as compared to NASC models (which are as highly endowed with physical meaning as possible, being actual samples of the plants themselves). To understand the empirical content of TAIR images, one needs to refer heavily both to GO categories and to the object-oriented approach to the organisation of data (as well as, in some cases, the circumstances and experiments yielding the evidence incorporated within each image). In this respect, TAIR images can rightly be claimed to be abstract, and certainly to be more abstract than NASC specimens.

The third sense in which a model can be abstract refers to the number of particular phenomena that it is taken to represent. Even more strongly than in the previous case, 'abstract' is here a relative notion whose application depends on the context: a model is the more abstract, the more specific situations it can be taken to represent.[169] This means that a description is abstract insofar as it can be 'fitted' to a number of other descriptions, which are therefore defined as 'less abstract' or 'more concrete' (Cartwright, 1999). According to Cartwright, something 'abstract' does not differ from something 'concrete' in an absolute, or qualitative, sense. The two notions are continuous and the characterisation of a model as abstract or concrete depends largely on its specific context of use.[170] In this third perspective, both NASC and TAIR models can be thought of as very concrete or very abstract, depending on the research goals of the scientists handling

[168] This is true of symbolic representations, such as a mathematical formula, for instance: it does not contain indications as to how it can be applied to actual phenomena. See Cat (2001) for an exploration of this notion of abstraction in relation to Maxwell's work.

[169] This sense of abstraction is similar to Radder's definition of abstraction as 'summarising', which he identifies (and goes on to criticise) as one of three main senses in which abstraction works (2006, 110). Note, however, that while his analysis applies to abstraction as a whole, I am here focusing on abstraction as used to produce models.

[170] It should be noted that Cartwright's discussion could be read as applying primarily to scientific discourse, and thus to symbolic or linguistic representations. This is because she does not explicitly consider the link between her notion of 'concrete', as applying to descriptions, and a notion of 'concrete' denoting a degree of materiality. However, even if Cartwright's account is supposed to apply to non-tangible entities such as concepts and sentences, the second intuition about abstraction creeps back into her examples of abstract versus concrete descriptions. Consider her classic example of the notion of work, which she takes to be an abstract notion with respect to the many different activities that it can be used to describe (such as cleaning dishes, marking exams, drilling a hole in the wall). There is a strong sense in which the sentence 'I am doing work' is abstract with respect to the sentence 'I am marking exams' insofar as it is largely devoid of physical meaning. The sentence 'I am marking exams' conveys a different type of information with respect to 'I am doing work': it indicates a material way in which I move, think and manipulate objects.

them. Prima facie, NASC models qualify as very concrete models, since they are only representative of the specific class of Arabidopsis ecotypes defined by their specific morphological and genetic make-up. However, many molecular biologists use a specific ecotype as a representative of any Arabidopsis variant - or even, in some cases, of any plant or living organism. In this latter case, NASC models are highly abstract. Similarly, the TAIR image of a metabolic pathway can be taken to represent any similar metabolic pathway found on any organism – thus 'fitting' a large amount of particulars. The same image can also be taken to represent a pathway typical of Arabidopsis plants, or even of a specific ecotype of Arabidopsis. In this latter case, the degree of abstraction characterising the TAIR model is lower than in the former.

These three intuitions about abstraction are often conflated in philosophical accounts of abstract models.[171] This is understandable, since all three contain a reference to the different meanings assumed by the term abstraction in common language (as well as by scientists themselves).[172] However, they spring from three very different philosophical concerns: the first, more ontological, with the extent to which entities are embodied; the second, epistemological, with the extent to which models contain information about how they apply to phenomena; and the third, also epistemological, with the extent to which models can generalise over and/or be used independently from particular contexts. As illustrated above, the difference in concerns underlying the three intuitions results in conflicting diagnoses of which models are abstract and in which ways and contexts. The conflation of these three notions of abstraction can prove uninformative, or even misleading, when uncritically used to characterise the differences in representational value among types of models.

I do not wish to argue that an analysis of model features as abstract and/or material would not be fruitful to a better understanding of their epistemological status, as well as of the value of the findings acquired through using those models. Yet, I am not interested in carrying out such an analysis here. Rather, I would like to propose a shift of perspective in the philosophical discussion of abstraction as characterising modeling practices: that is, from a view of abstraction as an attribute of models to a view of abstraction as an activity required for their production.

6.3 Skilful Abstracting

6.3.1 Abstracting to Model

I intend to address the debate on what it means for a model to be abstract by focusing on the processes required to produce a model. This implies a shift from thinking of

[171] See, for instance, Ankeny's ambiguous use of the notion of 'abstract entity' (2000, S264-6). Ankeny takes descriptive models of organisms (such as the wiring diagram representing C. elegans neurons) to be abstract in at least two of the senses suggested in my analysis. The wiring diagram model is: a. disembodied (it has no physical correspondent in any one specimen, since it is supposed to abstract over any specifically individual character in order to acquire generality) and b. widely representative (having broad representative value with respect not only to worms but to any simple metazoan nervous system).
[172] For a philosophical reflection on common-sense notions of abstraction, see again Radder (2006).

abstraction as a property of models ('abstract' as an attribute) to considering abstraction as an activity that is necessarily performed in order to produce models (the verb 'to abstract' as a way of acting) – a shift which, as I shall argue, throws light on the epistemic skills required to handle them in order to gain understanding of the natural world. I therefore wish to view abstraction as *the activity of selecting some features of a phenomenon P, as performed by an individual scientist within a specific context, in order to produce a model of (an aspect of) P*. In other words, I define the process of abstracting as involving the transformation of some features of a phenomenon into parameters used to model it, as relevant to the specific aspect of its biology that the model is deployed to study. For instance, in the case of Arabidopsis plants, abstracting means picking a limited set of properties of Arabidopsis wildtypes and use them as parameters to produce a specimens of Arabidopsis that always incorporates these features. A NASC specimen with these characteristics can then be employed to investigate and understand specific aspects of plant biology. In the case of TAIR images, instead, abstracting means selecting the aspects of the phenomenon that TAIR curators wish to model; interpreting the relations among those aspects in the light of a given theoretical framework (GO in this case); and producing a digital image containing symbolic representations of those aspects as well as suggesting how they might be connected with each other.

Defined in this way, abstraction is one of the processes required in creating a model, rather than an attribute of the model itself (the model thus being 'abstracted' in various ways depending on the specific circumstances and epistemic goals, rather than 'abstract' in an absolute sense). Further, it is an essential process in the context of modeling practice, as it is the process by which all types of models acquire a representational value with respect to some aspects of a phenomenon. As I shall illustrate, theory can play different roles in this process: the selection of features of P can be entirely based on theoretical assumptions (and thus make no direct reference to actual observations and interactions with P) or it can be largely independent of any theory, as it is based purely on a researcher's proto-explanatory exploration of P. I shall now illustrate how this applies to TAIR and NASC modeling practices. In the sections that follow I focus on the epistemological insights offered by this perspective on abstraction.

In the case of TAIR, abstracting is performed when selecting data resulting from an experiment in order to construct a two or three-dimensional visualisation of those data and the connections among them, relative to a specific component or process (such as a gene function or a metabolic pathway). NASC specimens are also abstracted from Arabidopsis wildtypes, insofar as they have to display specific sets of features of such wildtypes (such as a glabrous leaf surface), which are chosen on the basis of their relevance to research purposes as well as the extent to which they can be realised in a plant and replicated in others under controlled conditions. In both cases, then, abstracting is a crucial activity involved in modeling. Yet, it is carried out in markedly different ways. Abstracting as performed in TAIR implies selecting features of Arabidopsis genetics, biochemistry and cell biology by assessing their relevance to the interpretative framework that is (tentatively) used to study the entity or process under scrutiny. The main concerns underlying the production of TAIR images are their explanatory power, internal consistency (determined by reliance on the GO framework) and aesthetic value (simplicity and legibility), rather than the degree to which their parameters capture the relevant features of the plants. In other words, TAIR modellers prioritise an accurate rendition of the *relations* among elements of the model over an accurate rendition of the empirical meaning of these elements as represented. In visualising a metabolic pathway (figure 5), the emphasis is on the type of relation linking elements such as aminoacids and carbohydrates in order to enable metabolic processes. It is of secondary importance, within these models, whether the little triangles representing aminoacids and the little squares representing carbohydrates tell us something about the actual composition and structure of these substances as found in real plants. In this sense, abstracting towards TAIR images is largely an intellectual activity: it is theory-guided, geared towards explanation and requiring no physical contact with the phenomenal properties to be abstracted. I thus refer to it as *intellectual abstracting*.

In the case of NASC, abstracting is aimed at the material replication of features of the plants, whose stability across experiments and laboratory settings is a necessary condition to their representational value. The epistemic priority of NASC modellers is to maintain control over the development of traits characterising different ecotypes, thus ensuring the reproducibility of specimens as well as their non-locality (that is, the stability of their features regardless of the time and location of their use).[173] NASC researchers select specific features of the phenomena that they wish to model by reproducing these features across different generations of specimens: abstracted features of Arabidopsis plants are thus features characterising the morphology, physiology and genetic make-up of standardised specimens such as the Col (e.g. glabrous, round leaves, short stems, short life-span). This is realised mostly by modifying the growth environment and disposition of the plants: that is, by ensuring that they are sown and germinated in the best possible conditions (e.g. with enough space, water and humidity) and by growing them in isolated containers under artificially regulated light. Direct interventions in the plants themselves are also involved, in case they manifest undesirable or unexpected traits and when preparing the seed for storing and distribution. Here, abstracting does not imply establishing relations among given data about the plant by reference to specific theoretical frameworks (or, less generally, to actual explanations, as in the case of the explanatory powerful TAIR images). Abstracting involves focusing on the material features of the plants that need to be reproduced across specimens grown all over the world. This type of abstracting is performed by physical interaction between the researchers and the phenomenon to be modelled and is thus based largely on perceptually acquired knowledge about the phenomenon. While background theoretical knowledge is certainly involved (abstracting is theory-*informed*, as evident from the use of organisational concepts in PATO), it does not determine the activities and results of modeling (abstracting is not theory-*guided*). To emphasise the contrast with what I called intellectual abstracting, I refer to this type of activity as *material abstracting.*

A possible objection to this way of reading the processes involved in the production of material models is to question the label ‘abstraction’. Are these not processes of standardisation, that is, of constructing consensus around rules governing the production

[173] A detailed analysis of the notions of reproducibility and non-locality (in patterns as well as meanings) in the context of experimentation can be found in Radder (2003) and (1996, 2, 35-36).

of objects (Bowker and Star 1999, 13)? And if so, what is the use of referring to them as ways of abstracting? Some of the procedures that I list as part of the abstracting process are indeed determining standards (for instance, by establishing and enforcing criteria to define a plant as pertaining to a specific ecotype, or as being adequately grown in view of performing experiments). However, considering standardisation alone does not help in confronting questions about the epistemological value of material models. In other words, I am here interested in the production of representations rather than in the representation of production processes: thinking about abstraction in models means thinking about how models come to be representative of specific phenomena or issues, and with which consequences.

The distinction between intellectual and material abstracting immediately points to the different kinds of skills required to produce, handle and interpret models thus obtained. As clearly demonstrated in the cases of NASC and TAIR, modeling, no matter which types of activities it involves, is not carried out in a vacuum. Modeling activities are performed by particular individuals in a given context and with the help of specific instruments and tools, all of which factors impose material constraints on the way in which models are realised as well as on the type of biological understanding that can be achieved through handling them. As remarked in the previous chapter, researchers cope with the material constraints dictated by models themselves (as well as the context in which they are produced) by acquiring skills that enable them to handle models *correctly*, depending on the circumstances. The very adequacy of models as representations of natural phenomena is determined by the skill with which they are produced and handled as much as it depends on the features of models themselves.

As anticipated in Chapter 5, it is usually the epistemic community within which a scientist works that determines which skills should be exercised, and how, for a modeling activity to be judged as adequate. In producing NASC models, one needs to know how to handle plants, isolate them from each other, make sure they germinate and grow properly, regulate thermostats and ventilators and harvest seeds. TAIR modeling requires the ability to use a computer, type on a keyboard, write programmes in Java-script, search for relevant data by accessing specific databases on the internet or emailing/calling up the researchers responsible for them, and so forth. These are what I called *performative* skills. Then we have *theoretical* skills, like the ability to fit one's perception of a plant's morphology into PATO categories, or the application of the concept of 'gene product' to specific sets of data operated in TAIR. Further, there are *social* skills involved – such as the capacity to argue convincingly and intelligibly for the validity of a specific visualisation tool, as TAIR researchers do when presenting their work for evaluation at Arabidopsis conferences and meetings; or the co-ordination among technicians and researchers that is required for maintaining the NASC glasshouse and advertise NASC guidelines and distribution mechanisms to prospective users. I will say more about this last type of skills in the next chapter. The next two sections are dedicated to an analysis of the interaction of performative and theoretical skills in the two cases of intellectual abstracting and material abstracting.

6.3.2 Intellectual Abstracting and Theoretical Skills

Let us examine more closely the cases in which model manipulation and abstracting happens largely conceptually. In those cases, the main goals for the manipulation of the model are the testing, elaboration or illustration of a given theory about the phenomenon that is modelled. The goal is, in other words, to uncover ways in which a model can be *representative for* a given theory. The choice of the parameters used within the model is thus characteristically informed by a well-defined hypothesis about the theoretical outcome that the model is supposed to illustrate, test, predict and/or elaborate. This is because we start from a theoretically informed 'prepared description' of the phenomena under scrutiny. The term 'prepared description' was introduced by Cartwright, who defines it as 'presenting the phenomenon in a way that will bring it to the theory' (1983,133). A description of Arabidopsis microbiology constructed with the help of GO categories constitutes a good example for this. Importantly, Cartwright also argues that 'the check on correctness at this stage is not how well the facts known outside the theory are represented in the theory, but only how successful the ultimate mathematical treatment will be' (ibid.). Substitute 'mathematical' with 'conceptual' and this statement becomes applicable to the construction of TAIR images, in which, as we have seen, internal coherence and conceptual clarity have priority over the relations between models and the plants that they are meant to portray. Thus, according to Cartwright, the properties of the phenomena that are abstracted in order to be represented in the model, are properties that are either more causally relevant or more general (less context-dependent) than others: both generality and causal relevance are assessed in the light of a given theory.

The use of intellectually abstracted models is increasingly widespread among biologists. Take mathematical models, whose crucial role in the establishment of the 'Modern Synthesis' in the 1920s and 30s (Mayr and Provine, 1980), resulting in the birth of a whole discipline relying on statistical methods of analysis (i.e. population genetics), was only a prelude for their growing application across almost all biological disciplines. Even more evident is the pervasive use of simulations and algorithms to visualise empirical data, not to mention the push towards formalisation and away from the laboratory brought about by the increasing use of bioinformatics to store, organise and integrate data. These models are especially useful for elaborating explanations or confirming predictions stemming from given hypotheses (they are what Cartwright calls *interpretative* models in her 1999, p.181). They are also fundamental to the integration of biological knowledge concerning specific phenomena (as, for instance, bringing together insights from physiology, molecular biology, functional genomics and cell biology in order to understand root development in plants). However, precisely because of their strict reliance on theoretical assumptions, models constructed through intellectual abstracting are not the best of epistemic tools in cases where the goal of their manipulation is to improve the empirical content of a theory. They do not help with testing the empirical (descriptive) accuracy of the relation it stipulates between theoretical terms and aspects of the phenomenon.

This characterisation of intellectual abstracting does not leave doubt as to the importance of theoretical skills in order to successfully accomplish this activity. As we have seen in chapter 5, TAIR curators are devoted to creating models with high explanatory power. Accordingly, they exercise and value theoretical skills over and above the performative skills needed to handle the plants from which data are obtained. While working at TAIR, almost all curators are forced to forgo their experimental work and in fact they rarely acknowledge their expertise in handling plants as relevant to the construction of TAIR visualisations.[174] Their educational background is also geared towards theoretical skills, as the overwhelming majority of TAIR researchers have been trained in developmental biology (that is, the branch of life sciences that embraces most elements coming from other disciplines – such as physiology, evolutionary and molecular biology – and thus has sophisticated theoretical tools at its disposal for studying complex biological processes).

What about performative skills then? Skills relating to the handling of Arabidopsis plants are, as I already illustrated in 5.4, important to the interpretation of TAIR models and their use towards acquiring understanding of Arabidopsis biology. On the other hand, precisely because these skills play a minor role in the construction of TAIR images (and thus the abstracting of their properties), it is possible to handle these models without possessing this type of performative skills: only IT skills, involved in accessing and handling digital models, are required. Curiously, this latter type of performative skills is a prerogative of TAIR *programmers*, rather than TAIR curators. This sub-group within the TAIR research team specialises in bioinformatics rather than actual biology and thus lacks the theoretical skills required to making sense of the models that they develop. At the same time, their performative skills are crucial to the realisation of TAIR models. The division of labour between curators and programmers generates problems for TAIR users needing to handle TAIR images for their own purposes: to this aim, biologists need the theoretical skills used by TAIR curators (and thus need to be acquainted with the GO framework) *as well as* the performative skills characterising the work of TAIR programmers (without which it is difficult to gain access to models to start with, not to mention to manipulate them as generally required by researchers).

The dominant role of theoretical skills, as well as the relegation of a great part of performative skills to technicians rather than biologists, shows how TAIR models are largely intellectually abstracted: their features are selected on the basis of a given theoretical framework (the GO) and in response to the research interests expressed by the prospective users of those images. The emphasis on producing models with high explanatory power does, however, diminish their empirical content, as the increased theoretical significance is gained at the expense of materially abstracted features within the model. This is a typical trade-off situation among the various epistemic values that models can possess, as described by Levins and mentioned in chapter 2. I am now in a position to explain the necessity of a trade-off between explanatory power and empirical content of models in terms of the skills involved in abstracting them. In the case of

[174] This does not mean that they do not recognise their embodied knowledge, as acquired through experimental work, as crucial to accomplishing their tasks (as I discussed in section 5.2.3). Rather, it indicates that TAIR curators bestow little attention to this aspect of their work: they recognise it and are able to discuss it, but only when they are explicitly questioned about it.

intellectually abstracted models such as TAIR images, the performative skills required do not concern the handling of plants themselves (thus, the interaction between biologists and phenomena): rather, they concern the building of images that conform to the 'vision' devised by TAIR researchers though reference to their theoretical skills. Theoretical skills thus play a primary role in the construction of TAIR images, to which performative (IT) skills are subsidiary: the activity of abstracting is largely intellectual, thus generating models with high explanatory power but low in empirical content. Let us now see how materially abstracted models can complement intellectually abstracted ones, namely by adding empirical content to such theory-rich representations.

6.3.3 Material abstracting and Performative Skills

When thinking about the type of description needed to model a phenomenon, Cartwright proposes to focus also on 'unprepared descriptions'. These are descriptions that (i) 'contain any information we think relevant, in whatever form we have available' and (ii) 'are chosen solely on the grounds of being empirically adequate' (1983, 133). As an example, consider the free-text descriptions of Arabidopsis morphology provided to NASC by donors of specimens. These descriptions provide a lot of empirically adequate information, yet are not phrased in ways useful to NASC researchers in order to grow and categorise models. To that aim, the descriptions have to be 'prepared', i.e. modified and standardised through PATO.

I consider Cartwright's account of unprepared descriptions as a fruitful recognition (contra more traditional accounts of modeling that insist on referring to theoretical physics as a 'role model' for all other sciences) that the testing of theories need not be the starting point of investigation in experimental science. What we witness in experimental biology is the use of several types of models that are not built to verify or extract predictions from a given theory. Their use is unavoidably guided by background knowledge and a commitment to the investigation of specific conceptual issues. Yet, the background knowledge needed to formulate a question should not be confused with the knowledge produced by trying to answer it. In preparing a description, scientists rely on an already formed hypothesis about how to answer a given theoretical question. The manipulation of unprepared descriptions, by contrast, does not require the choice of a specific interpretation to start with. That is to say that the manipulation of models sometimes requires no more than a general interest in exploring one or more aspects of the phenomena that they are taken to represent. In fact, and here I part company from Cartwright's 1983 account and I move into her revised 1999 framework, there are cases in which the unprepared description of a model is constituted by diagrams, objects or even samples of the phenomenon itself. NASC specimens themselves are a case in point: these organisms are taken to be *representative of* a set of phenomena (amounting, depending on the research context, to 'all plants', 'all flowering plants', 'all weeds' or 'all other Arabidopsis ecotypes'), in the sense of being used to explore which properties of the phenomenon could turn out to be relevant to a theoretical account of it. Thus the model provides the epistemic access to phenomena that is necessary in the first place, in order to infer the kind of unprepared description that Cartwright is talking about.[175] Epistemic access is granted first and foremost by material manipulation and abstracting, since the amount of intellectual manipulation and abstracting necessary to handle these models is minimal. The theoretical framework that they are representative for does not need to be specified in order for scientists to use them, since it will eventually be developed through other, complementary forms of modeling.

Contrary to my characterisation of models produced through mostly intellectual abstracting, models derived through material abstracting belong to a ‘proto-explanatory context’ (Ankeny 2001). In such a context, scientists not only have not agreed on a theoretical explanation of the phenomena under investigation, but have not even settled, yet, on which properties of those phenomena could be relevant to the explanation (a decision that is crucial to the building first of an unprepared, then of a prepared description, thus enabling the shift from largely material to largely conceptual manipulation - and back). Such models are often tangible objects like scale models, samples, robots or, most emblematically, model organisms, which the scientists interact with through their sense perception.[176] As illustrated in section 6.2, the ambivalent nature of materially abstracted models such as NASC specimens constitutes one crucial reason for their effectiveness as exploratory tools. Griesemer stresses this aspect when declaring that 'material models are able to serve certain sorts of theoretical functions more easily than abstract formal ones in virtue of their material link to the phenomena under scientific investigation. [..] They are robust to some changes of theoretical perspective because they are literally embodiments of phenomena' (1990, 80). It is the degree to which materially abstracted models represent phenomena independently of a theory that makes them, in my view, take a substantially different epistemological role with respect to intellectually abstracted models as characterised above. This is illustrated by the performative skills refined by researchers in order to abstract their material features.

To NASC modellers engaged in material abstracting, performative skills are essential to abstracting reproducible features of the plants, as is evident from the crucial step 2 of the procedure (‘*cultivation of specimens*’). NASC researchers select which aspects of Arabidopsis plants should be abstracted largely by experimenting with the plants and seeing which of their features can be reproduced in a stable and predictable way under laboratory conditions. The features of the plant that are abstracted within NASC specimens are those over which researchers hold some degree of control: the features that, through specific performative skills acquired by iterative handlings of the plants, can be reproduced across generations of specimens. A case in point is the skill, exemplified in

[175] Cartwright herself recognised this in her most recent work and thus distinguished these ‘representative models’ from interpretive ones (1999, 180). Morgan also emphasises the explorative role played by models in her (2003) and (2005).
[176] A scale model or an organism might seem far too complex an object to play a representational role: yet interacting with it - feeling, choosing, discarding and comparing its characteristics (whether implicitly or explicitly) – often leads scientists to consider alternative explanatory frameworks, precisely because of the dynamic nature of their features. On this point, see also Giere (1999), Magnani (2001) and Polanyi (1958).

section 6.1.3, to recognise plant ecotypes and discard those whose characteristics do not conform to the expectations of researchers (figure 6.5).[177]

The fact that performative skills represent a primary means to perform material abstraction is reflected in the background and skills of researchers hired to produce NASC models. Researchers working in the NASC laboratory and glasshouses need an educational background as technicians and/or experimentalists. NASC director Sean May values researchers with good material skills and a practice-oriented approach much more highly than he does appreciate theory-directed scientists whose focus is on explaining, rather than acquiring, data (such as the ones working at TAIR; *pers. com*.). There is relatively little place for theoretical skills in the abstracting of NASC specimens: they are exploratory tools, whose features are not chosen because they fit a specific theoretical framework or skill, but rather because they are easily transformable through mutation and thus can be used to verify the effects of various kinds of experimental interventions in the plants (this is certainly the case, for instance, with the Lan ecotype of Arabidopsis). Researchers who need to handle specimens need first and foremost the relevant performative skills, as acquired in their own laboratory experience or, in part, through protocols and guidelines about how to handle plants divulgated by NASC. PATO descriptions might be viewed as bringing some theoretical input into the abstraction process. Their use in the making of NASC specimens requires theoretical skills such as acquaintance with bio-ontologies, knowledge of the PATO descriptions that apply to the description at hand and the ability to restructure the donor's description along PATO lines. Yet, these theoretical skills are as subsidiary to performative skills characterising NASC practice as the performative skills employed in the abstracting of TAIR models are subsidiary to the theoretical skills of TAIR curators.

Performative skills are part of the embodied knowledge gathered by biologists who conduct actual experiments on Arabidopsis plants: they are one reason why material experiments on organisms are very difficult (if not impossible) to substitute with virtual experiments performed on digital reproductions of those organisms – the latter being far less efficient than the former in enabling biologists to coordinate their theoretical knowledge with embodied knowledge of the phenomena under scrutiny. Current attempts to carry out biological research on purely theoretical, or even digital, platforms (such as TAIR) are bound to provide a limited understanding of biological phenomena, precisely because these activities do not help acquiring the performative skills needed to make empirical sense of the theoretical knowledge therein contained. Biologists need to intervene in organisms in order to understand their functioning. This is true of Arabidopsis plants as, I expect, it is true of fruit flies, mice, rats and even monkeys. This remark is not meant to justify all experimental practices on animals and plants. I agree with animal right activists that much of this research is futile, since it is often carried out for scientific goals of little import (such as skin-tests commissioned by the cosmetic industry), as well as unnecessarily cruel, insofar as it shows little concern for the animals' welfare (animals are kept in deplorable conditions and are often made to suffer far more

[177] I here follow Radder (2006) in suggesting that observation is an epistemic activity to be performed skilfully, as there are different ways in which a scientist can observe and different conclusions can be drawn from the same observation.

than strictly necessary for achieving the given research goals). However, I do want to point to the irreplaceability of such research in biology. Unfortunately for the animals, there is no way in which the use of model organisms as models in biological experiments can be replaced by the use of other types of models.[178]

6.3.4 Skills in Abstracting: Indispensable and Inseparable

I have been arguing that, while noting the different degrees of materiality or abstraction embodied by NASC and TAIR models is not particularly informative towards clarifying their role in acquiring understanding of phenomena, the distinction between conceptual and material abstracting of models allows us to note the different skills required in order to handle them. Experimentation on actual plants allows biologists to learn things that they would not learn by consulting TAIR or running a few simulations of plant development. They do not only acquire performative skills: they also learn what the plants are like, what are their structure, texture and responses to specific handlings, what is their response to the environment and specific parasites. All of this knowledge, which I defined as embodied knowledge, informs and complements the performative skills acquired in order to handle the plant. Through exercise of these skills, biologists use their embodied knowledge of a phenomenon to intervene in the phenomenon itself.

As is evident from my analysis so far, the gap in values, training and skills privileged within the NASC and TAIR research teams is so strong as to generate intellectual and social tensions between them.[179] This reflects the striking difference between the types of skills accompanying different types of abstraction processes. At the same time, the distinction between material and intellectual abstracting does not in any way preclude recourse to performative skills to perform intellectual abstraction, nor does it preclude recourse to theoretical skills to obtain material abstraction. Abstracting as operated in TAIR can be said to involve performative skills, insofar as it requires researchers to physically fumble around alternative visualisations of the same set of data and categories (as I pointed out, TAIR modeling involves manual curation to a large extent).[180] Abstracting towards NASC models also implies several theoretical skills: the way in which specific traits of plants are categorised is partially conceptual, as exemplified by the reference to PATO networks of concepts (which are themselves, however, much less

[178] Note that questions about the extent to which animals or plants can be representatives for other species – especially humans – is not confronted here. It constitutes an entirely different issue, which I have no interest in tackling within this context. What I want to discuss here is not whether the representative value of model organisms, as attributed by the research community using them, is justified. Rather, I focus on the conditions under which model organisms become models and thus contribute to biologists' understanding of phenomena that they are assumed to represent.

[179] Both groups argue that their approach to modeling is better than the other's, an attitude that does not – thankfully - entirely prevent their collaboration, but that signals the extent to which their modeling strategies differ. These tensions have become especially clear to me through conversation with Sue Rhee, the director of TAIR, and Sean May, the director of NASC. More on this point in Chapter 7.

[180] A relevant difference between manual curation of NASC specimens and TAIR images is, however, that NASC specimens are, at least in part, samples of the very phenomenon that they are meant to represent – thus, material abstracting involves manipulation of the phenomenon to be abstracted as well as of the model, while intellectual abstracting does not involve handling the phenomenon to be represented.

theoretical than the GO framework, since they are constructed in order to describe the morphology of specimens – 'glabrous', 'leaf' - rather than in order to capture the meaning of notions used to describe processes, such as 'metabolism' or 'pathogeny'). Indeed, the intertwining of performative and interpretive skills to perform any type of abstracting is unavoidable: as demonstrated by recent results in cognitive science, even the most conceptual and rationalistic human activity always implies physical interaction with the world. All human actions, including the ones aimed at producing scientific knowledge, are necessarily both embodied and reflexive. This means a constant interplay between theoretical and performative skills, which enables researchers to coordinate their theoretical knowledge with their embodied knowledge in order to understand a specific phenomenon. The example of TAIR and NASC models in Arabidopsis research usefully illustrates this point. The role played by TAIR models towards their users' understanding of Arabidopsis biology is greatly enhanced by these users' experience with largely materially manipulated models, such as NASC models.[181] That both TAIR and NASC recognise this is evident from their agreement to make NASC databases searchable in TAIR, as well as from TAIR recommendations to its users to order the NASC seed of the recommended variety in order to test or further an existing set of data.

6.4 Modeling to Understand

The main goal of this chapter has been to illustrate, by comparing activities carried out at TAIR and NASC, how the coupling of theoretical and embodied knowledge achieved through modeling practices might bring about an understanding of the biological phenomena being modeled. Modeling activities are a striking case of this as well as an important one, given the prominence of such practices in biological research. The focus on the epistemic skills required when abstracting and handling models has hopefully highlighted the importance of embodied knowledge to the acquisition of a biological understanding of phenomena. As I argued in Chapter 2, theoretical knowledge needs to be complemented by embodied knowledge in order for researchers to make sense of the aspects of the natural world that they are studying. The ability to manipulate and, to a certain degree, control a phenomenon is essential to acquiring biological understanding: without this ability, biologists would not be able to single out features of a phenomenon

[181] My claims about the necessity of theoretical and embodied skills in order for users to make sense of TAIR categories and datasets, which I also emphasised in Chapter 5 when addressing the embodied knowledge involved in interpreting TAIR results, run against an argument proposed by Anne Beaulieu (2001) in her study of the role of digital modeling and database construction within the neurosciences (which, as she recounts, bears remarkable similarities to digital resources and informatics in the biomedical sciences). Beaulieu insists on the valuable features of integration through digital tools, thus buying into the idea (much publicised by curators directly involved in making these tools) that 'neuroinformatics foster large-scale sampling and automated processing – remote from the individual as embodied object or subject' (2001, 661). Beaulieu's fascination with these practices leads her to propose the digital manipulation and integration of data as a path towards 'a new digital objectivity' (2001, 664). As evident from my analysis so far, my view on the value of bioinformatics is not as rosy: while also appreciative of the advantages provided by the new digital tools, my study of the theoretical and performative skills involved in using these tools shows that researchers using them always require personal input and interpretation. Their contribution to integration in the biomedical sciences can thus hardly be described as guaranteeing objectivity through automation.

that are relevant to explaining its behaviour; to apply their theoretical concepts to the observations that they make; or to coordinate their thoughts and actions in the reproducible way that characterises scientific understanding and makes it one of the most interesting results of human cognitive abilities.

Scientific understanding requires both interaction with the phenomena to be understood and reflection upon their features: performative skills play the crucial role of granting access to various features of the phenomenon and determine which ones can be subjected to experimental studies; theoretical skills allow to reflect on those features, classify them on the basis of a theoretical hypothesis or framework, and determine inferences from those features to an explanation of the functioning/behaviour/structure of the phenomenon at hand. In short, biologists value knowledge that they can efficiently use in order to understand phenomena. Skilful modeling does not only require, but actually enhance researchers' ability to use theories, protocols and other types of theoretical and embodied knowledge towards the understanding of a specific phenomenon. This constitutes the essence of what I called 'coordination of knowledge' in Chapter 2. In the same chapter, I also indicated how, in the case of understanding through modeling, such coordination needs to be both skilful and efficient: 'skilful', in the sense hitherto discussed of requiring expertise; 'efficient', in the sense of allowing researchers to understand the phenomenon that they intended to study in ways that are coherent with their previous knowledge and understanding of related phenomena (and, thus, coherent with the research goals that they set out to accomplish). My analysis of TAIR and NASC modeling practices has amply illustrated how different types of modeling activities require different combinations of skills in order to be adequately performed. As I argued in Chapter 5, such skills are indispensable to acquiring and coordinating the knowledge (both theoretical and embodied) that grants understanding of the phenomena being modeled. Here is thus a first important link between modeling and understanding. The use of a variety of models requires researchers to learn and exercise several theoretical and performative skills. These skills are potentially helpful to the pursuit and acquisition of biological understanding, depending on the specific phenomena to be understood as well as on the kind of understanding sought by researchers.[182]

I now wish to highlight another way in which epistemic activities involved in constructing and handling models crucially contribute to a researcher's understanding of a biological phenomenon: modeling provides grounds on which to identify knowledge that is *relevant* towards the understanding of a given phenomenon. This is a very important feature, as the knowledge of the natural world possessed by researchers trained for scientific inquiry is usually extensive: finding which parts of such knowledge – whether theoretical or embodied – are relevant to understanding the specific phenomenon under scrutiny is one of the most difficult tasks to be accomplished by biologists. Relevance can be established according to two main criteria: (1) the extent to which that knowledge fits the properties of the phenomena at hand (and allows researchers to manipulate them, whether conceptually or materially); and (2) the extent to which that knowledge serves the research goals and interests pursued by researchers.

---

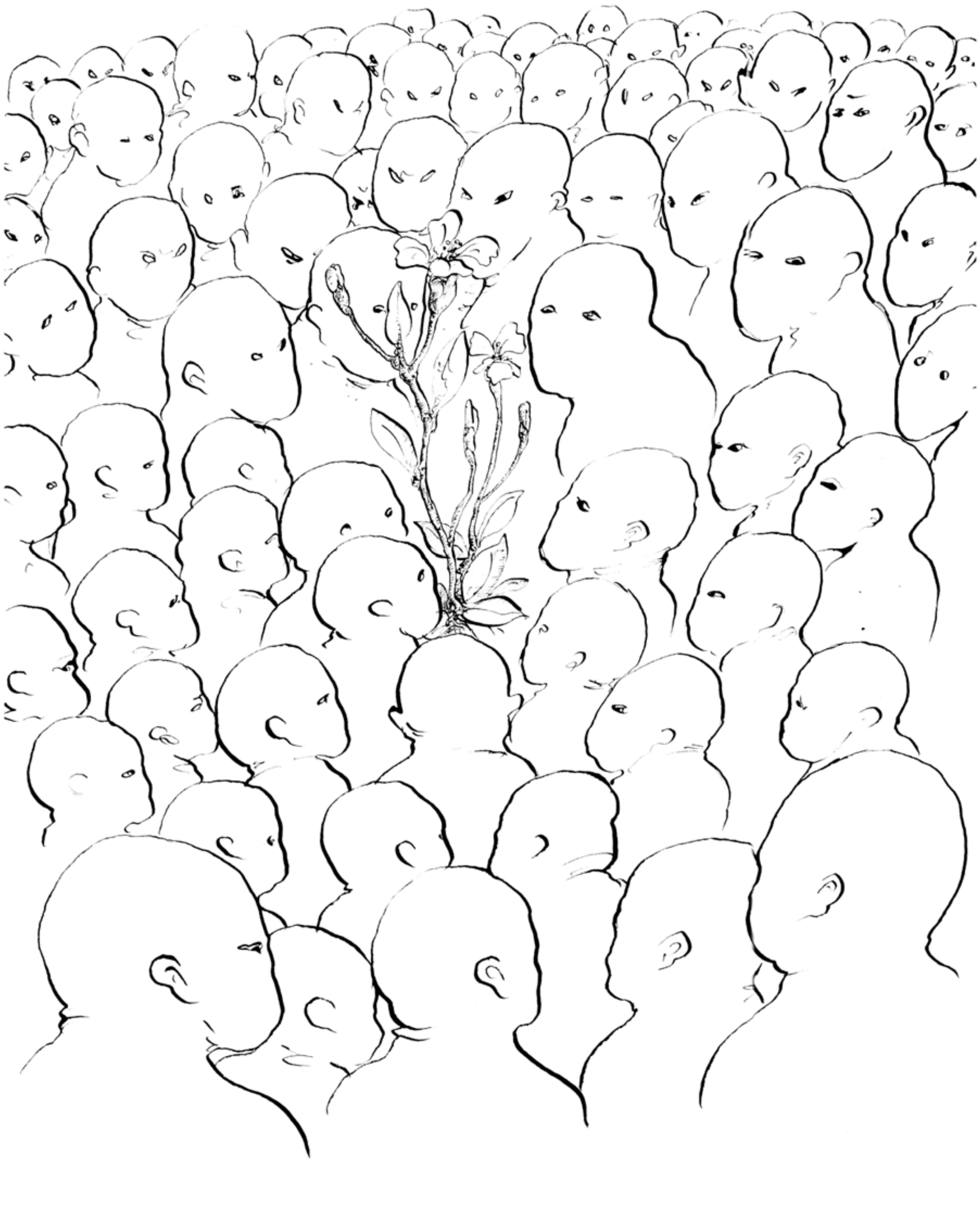
[182] More on kinds of understanding in Chapter 8, section 8.3.

Concerning (1), I have discussed how modeling via skilful abstracting constitutes a particularly efficient way to match properties of a phenomenon with specific parts of the theoretical knowledge possessed by a researcher. Selecting the features of the phenomenon that are to be abstracted involves selecting, at the same time, theoretical and embodied tools through which these features can be identified. For instance, modeling a metabolic pathway means matching data about the molecular structure of nutrients in a cell with terms that describe and classify those structures (such as amino acids, proteins, carbohydrates and so forth). The visual representations of metabolic pathways produced by TAIR allow to check and improve on the fit between theoretical knowledge and empirical data, In this sense, modeling enables researchers to fit data to interpretations, as well as to use those interpretations to guide their interventions on phenomena. Further, modeling allows researchers to distinguish knowledge that they wish to hold as an undisputed premise of their research (so-called ‘background knowledge’) from knowledge that they are prepared to test, challenge or modify as a consequence of their findings. The construction of GO is a good example of this: by building GO, TAIR curators have effectively established the theoretical boundaries within which Arabidopsis researchers, especially those working on molecular biology, ought to work. The importance of distinguishing knowledge used as a premise from knowledge that is to be tested was evident in my analysis of material abstracting, where I pointed to the use of theoretical knowledge as a platform from which to ask questions, rather than as a resource with which to formulate tentative answers. In a similar way, the goal of intellectual abstracting is not to acquire new embodied knowledge or challenge the performative skills already gathered, but rather to use that embodied knowledge as a basis for pursuing theoretical hypotheses and thus further existing theoretical knowledge. Without basic and undisputed assumptions, beliefs and habits (many of which are acquired through socialisation and training, as I shall discuss in my next chapter), it would be impossible to attribute a *representational value* to models: what a model stands for depends both on its features (which constrain the manner in which the model can be usefully manipulated) and on the research context in which it is used. Similarly, *which* aspects of a phenomenon are understood (and *how*) is determined as much from the modeling of the phenomenon as from the way in which researchers formulate their questions about the phenomenon (which in turn depends on their research goals and background knowledge).[183]

Let us now turn to the second criterion for relevance, that is, the necessity for each researcher working in a specific community to adhere, at least in part, to the epistemic culture of that community, including its interests, goals and values. The knowledge acquired and used by an individual scientist to research a phenomenon is extremely specialised and highly conditioned towards the areas of knowledge already selected as

[183] In Chapter 8 I shall come back to these points and their significance towards an analysis of understanding. In particular, I shall focus on the *commitments* under which research is being carried out. In section 5.5 I argued that the same phenomenon can be understood in different ways by researchers equipped with different combinations of epistemic skills. In fact, such pluralism depends not only on skills, but also on the different commitments to specific concepts, gestures, theories and perspectives that skilful actions such as the ones involved in modeling might bring about. ‘Relevant’ knowledge identified through skilful modeling can actually become the object of such a commitment, if it becomes entrenched in scientific inquiry to such an extent that it is not anymore subject to extensive critical scrutiny.

relevant by the community within which such research is carried out. This means that the interests, goals and values characterising that community play a role in selecting the knowledge that is to be considered of relevance to understanding a specific phenomenon. In fact, the very skills used to model phenomena are formed through appeal to knowledge, interests and values that transcend individual experience. They are shaped by participation in one or more scientific communities: the ‘right’ type of understanding, that is the one fulfilling the expectations, goals and questions posed when starting research on a phenomenon, is determined as much from individual researchers’ experiences in handling and modeling that phenomenon as it is shaped by the social context within which research is conducted and evaluated. In the next chapter, I shall therefore turn to the relevance of the social dimension to the acquisition of biological understanding.



Chapter 7. Community Matters

> *In a very real sense, we create knowledge when we give it to more people. And the acquisition of the 'same' piece of knowledge by every new person will have a distinct meaning and import within that individual's system of beliefs. When it comes to knowledge, dissemination is a genuine form of creation*
>
> *Hasok Chang 2004, 243*

In the previous two chapters, I focused on how an individual biologist might understand a phenomenon by skilfully coordinating theoretical and embodied knowledge as relevant to this aim. Yet, as I discussed in section 2.3.1, the skilful and efficient manner by which such understanding is obtained is not enough to qualify it as scientific (or, more specifically, biological). In order for understanding to be recognised as scientific, the biologists acquiring it need to be able to communicate their insight to their peers, so as to make it vulnerable to public scrutiny and evaluation. Individual understanding becomes scientific only when it is shared with others, thus contributing to the growth of scientific knowledge and partaking in the rules, values and goals characterising scientific research.[184]

Of course, the interpersonal communication of biological understanding looks like a paradoxical requirement, given my definition of such understanding as the *cognitive* achievement of an *individual* scientist, based on *his or her own* experiences, skills and knowledge. And in fact, I do not wish to argue that individuals can disseminate their understanding of a phenomenon in an unmediated, direct way – that is, for instance, by talking to other individuals. I want to propose that individuals possessing a specific understanding of a phenomenon might enhance other individuals' chances of acquiring it in an indirect manner: that is, by constructing tools (including, for instance, models, explanations, experimental set-ups and materials) that will enable other individuals to learn and exercise skills in ways that might eventually lead them to experience the same kind of understanding. This type of indirect sharing of understanding characterises both the acquisition and the dissemination of understanding in scientific communities. On the one hand, biologists seeking an understanding of a phenomenon are required to learn as much as possible from similar efforts by other scientists, so as to use the understanding accumulated by others as a vantage point for starting their own research. On the other hand, biologists who already understand a specific phenomenon are required to contribute to the body of theoretical and embodied knowledge available on that phenomenon in ways that will help other scientists acquiring the same insight. I argue that the social processes through which biological understanding is acquired and disseminated have a strong impact on the features of such understanding, One way to analyse such impact is

[184] Other types of understanding besides scientific understanding might also require dissemination and/or discussion among different individuals. What I want to stress here is that, in order to be classified as scientific, understanding needs to be shared within a scientific community. Only under this condition does understanding a phenomenon count as scientific, no matter how interesting or useful it might be otherwise.

to focus on a third category of epistemic skills that is involved in the acquisition of scientific understanding, that is the category of *social skills*.

This chapter is dedicated to a reflection on social skills and their role towards the acquisition of understanding. In the case of Arabidopsis research, such reflection involves an analysis of the ethos and structure of the Arabidopsis community as well as of the relations of epistemic dependence existing among individuals and/or groups within and outside the community (7.1). This analysis of the social organisation of Arabidopsis research, which I identify as a case of centralised big science, will bring me to list the social skills without which it would be impossible, for individual members of the community, to understand aspects of Arabidopsis biology in a way that can be shared by other members of the same community (7.2). This is, on the one hand, because individuals need to use social skills in order to be trained and constantly updated on the theoretical and embodied knowledge accumulated in the community: as I explain in section 7.2.1, exposure and participation to one or more scientific communities provides researchers with at least some of the theoretical and performative skills that they require in order to understand the aspect(s) of Arabidopsis biology that they are interested in. On the other hand, the exercise of social skills is paramount also in order to share individual understanding with the rest of the community, thus making it open to constructive criticism and eventual improvement (and therefore transforming it into a contribution to the growth of the established body of scientific knowledge). As I illustrate in section 7.2.2, it is possible for a researcher to learn to build theories, models, explanations and observations so as to make it easier for other individuals to understand a phenomenon in the same way as he or she did.

The significance of this series of claims will be discussed in the last two sections of this chapter, where I shall use my view to evaluate a philosophical view previously proposed to make sense of the constraints posed by community dynamics to the making of scientific knowledge. This is Helen Longino's account of the norms that a community needs to follow to produce what she calls 'epistemically acceptable' knowledge (7.3). I shall conclude that, while providing a rich and informative perspective on the relevance of socialisation to scientific knowledge, Longino's disregard for the issues relating to understanding leads her to omit some crucial elements, which results in a weakening of her whole approach. Through this example, I hope to illustrate how my analysis of scientific understanding can illuminate new aspects of some long-standing debates in the philosophy of science, especially within the area of social epistemology.

7.1 Epistemic Dependences

Dear Aunty Raby,

I am 24 days old and really desperate. I still haven't flowered and my rosette leaves are beginning to die. Do you think it's my appearance? Why do maize plants have all the fun? Please help me before I introgress myself.

Erica Erecta

Dear Erica,

Don't worry! See a specialist. I recommend genetic counselling. You may have an inherited mutation. Why not try 10µm gibberellic acid, it's a wonderful pick-me-up. I always find it helps.

Aunty Raby

'Agony Column' from Rabido, the Fifth Arabidopsis newsletter, July 1991

The international newsletters circulating around the Arabidopsis community at the end of the 1980s, when communication among members was still dependent largely on paper rather than on digital tools like email and the internet, represent an excellent source for the historian trying to reconstruct the early history of research on this organism. They are rich of information concerning the groups participating in Arabidopsis research, the increasing coordination of their research efforts and the astonishing growth of results harvested from studying this plant. A particularly striking feature of these newsletters, which testifies to the friendly atmosphere in which the Arabidopsis community developed (at least in the United States and the United Kingdom), is the wealth of extra-scientific material with which they are filled: jokes, cartoons and caricatures (such as the 'Agony Column' reported above) abound among those pages, with researchers contributing even some (rather atrocious) poetry and recipes centred on Arabidopsis. That so much enthusiasm, energy and creativity was invested in the writing of a newsletter – that is, a means of communication of no academic prestige whatsoever[185] - is a telling signal of the importance of inter-personal ties and shared values in the development of Arabidopsis research.

An even more significant factor pointing to the uncommonly friendly ethos characterising the Arabidopsis community is the ease and openness with which researchers used the newsletters as an informal platform for exchanging results and insights at the pre-publication stage. This is an exceptional occurrence, especially given that most Arabidopsis researchers involved in the community at the time were molecular biologists. As narrated in chapter 3, molecular biology has been a very competitive field for over four decades now: researchers working on other organisms usually refrain from sharing their results for fear that other researchers might steal and publish them beforehand, thus gaining all recognition. This competitive system is enhanced by the tendency by sponsors to consider the number of publications issued by any one laboratory as an absolute measure of the quality of research conducted there: funding is given to the groups that publish the most, thus making researchers even more defensive of their results and cautious of the possibility that others might steal them or publish similar data beforehand. Such caution engenders a lack of collaboration among biologists and especially among different research groups working on similar phenomena. This is damaging to the quality

[185] Articles published in newsletters, as well as all the rest of the work needed to organise and publish them, are not usually recognised as a valuable component of a researcher's CV: in fact, they are mostly seen as distractions from conducting 'real' research and preparing more valuable publications. This is despite the enormous importance of these means of communication to enhance the internal cohesion and information flow within a research community (a function well recognised within the Arabidopsis community).

of the research produced, while also enhancing the risk of different groups working on the same phenomena with the same objectives and tools without being aware of it – thus wasting resources and time that could be spent on other, more fruitful projects. The ethos of the Arabidopsis community, which TAIR director Sue Rhee aptly characterises through the motto 'share and survive', constitutes a surprising exception in this context.[186]

In fact, allegiance to this surprisingly cooperative ethos has played an important role towards transforming the Arabidopsis community from a little group of like-minded, young researchers to one of the largest cases of centralised big science in contemporary biology. As shown in chapter 3, the close personal ties among the founders of the community, as well as their wish to enforce and maintain an ethos of openness and collaboration among researchers working on Arabidopsis, were crucial to the success of Arabidopsis as a model organism and to its adoption by so many laboratories around the world. Those research groups were not only attracted by evidence showing the advantages of using Arabidopsis, rather than other plants, for experimentation: it is not so clear that those advantages were decisive in choosing a species of weed over plants of more obvious agronomic interest (such as yeast, rice or tomatoes). A strong incentive for new participants to Arabidopsis research was the common ethos enforced across the community, its highly centralised structure and the ease with which it seemed to attract funding (three factors which I referred to, respectively, as 'social commitment', 'institutional organisation' and 'funding sources' in section 3.2.3). It is therefore on these three characteristics that I would like to focus now, placing particular attention on the types of epistemic dependence that they enforce in the Arabidopsis community (and thus, their impact on the tools and interests employed by researchers in their quest to understand Arabidopsis biology).

7.1.1 An Ethos of Accessibility

Each one of the 5000 laboratories working on the model tackles questions concerning Arabidopsis biology with a unique and specialised combination of skills and expertise. Thus it is impossible to define Arabidopsis research as centring around a specific methodology, a unified research programme or a disciplinary approach. Even the instruments used in the community tend to vary. For instance, researchers working within the subdiscipline of molecular biology employ tools ranging from microarrays and affimetrix (i.e. sophisticated and expensive new techniques allowing to track gene activity across whole gene clusters) to classic sequencing and chromosome analysis. Surely, reference to the same object – Arabidopsis plants – is the main factor bringing all

---

[186] It could be argued that the peculiar nature of the Arabidopsis community ethos greatly diminishes its representativeness with respect to other model organism communities in biology, thus weakening the scope and applicability of my philosophical analysis. Of course, it is true that the social organisation of the Arabidopsis community is more of an exception than the rule within biological research. I am convinced that, for the purposes of my analyses, this is actually an advantage rather than a disadvantage: the exceptional nature of Arabidopsis research allows me to highlight issues (such as the tension between centre and periphery or the difficulties involved in interdisciplinary collaboration) that characterise most model organism research, but that are less visible in other communities.

these different approaches and perspectives together. Yet, what makes Arabidopsis researchers into a coherent community is their constant communication and data-exchange, as well as their allegiance to a set of common research goals and institutions. The communication patterns and ethos of social cohesion initiated within the community since its foundation are largely responsible both for its success and for the initial choice of the plant; further, the construction of integrating models such as NASC specimens or TAIR images would be impossible in the absence of the network of institutionalised collaborations that devised and supported both the NASC and the TAIR projects.

Since the early 1980s, when Arabidopsis was re-discovered as a model organism and research efforts on it acquired momentum, an agreement was stipulated among some prominent Arabidopsis researchers that the data acquired through research on this model organism would be kept freely available.[187] In this respect, the Arabidopsis community is unique among the main model organism communities. In no other case have the modalities and division of labour in studying the organism been 'rationally planned' from the very set-up of research. Both biologists and research institutions involved in Arabidopsis research invested and continue to invest considerable effort in trying to ensure that such agreement is respected as much as possible, especially given the increasing commercialisation of basic biological research (prominently in the form of patenting, today a diffused practice in virtually all model organism communities). The agreement on free exchange makes it possible to build and maintain a constant stream of communication among researchers, both on a personal level (through regular international meetings and conferences) and in publishing (through peer-reviewed papers and centralised online databases, which are now substituting the international newsletters in use from 1964 until 1994). It also ensures that Arabidopsis research is largely conducted within academic institutions rather than corporations (though, as I shall discuss in the next sections, this does not imply that there is or could be a strict distinction between basic and applied research in the community, nor that industry did not influence the research under way).

It is tempting to relate the adoption of this ethos and organisational structure to the extraordinary scientific success of the Arabidopsis community, which, in little more than two decades, has unravelled some of the most important recent discoveries in molecular and developmental biology (including studies of the role of hormones in development, methylation processes and the role of various kinds of RNA in the expression of genetic material). It is not obvious, nor can it be proven, that observance of such an 'ethos of accessibility' has been one of the main causes for the great increase in the understanding of plant biology acquired by researchers studying Arabidopsis. Yet, causal ties between the two factors are very likely. Observance of this ethos ensured that power struggles and competitive tensions did not entirely prevent exchanges of ideas, data, instruments and materials between researchers. Biologists involved in the study of Arabidopsis do not only share their interest in this specific organism (which would, on its own, already warrant the attempt to share as much information as possible, in order to rationally distribute research efforts and thus avoid duplicating results). They share some broad goals and beliefs, among which the assumption that Arabidopsis can be representative for

[187] For details on this agreement, see Chapter 3, section 3.1.1.

many (if not all) other plants and the belief that it is important to try to integrate knowledge about different aspects of Arabidopsis biology. Most important for the purposes of maintaining a common ethos, many Arabidopsis researchers share the idea that a joint effort to integrate molecular biology with other disciplines may lead to the re-invention of plant biology as a whole, thus transforming it from a field consisting largely of taxonomy and natural history (a feature that contributed to its becoming highly unpopular with respect to 'intellectually exciting' animal biology) to a field whose understanding of organismic development, growth, morphology and ecology is both descriptive and mechanistic - that is, whose advances are informed by information about processes happening at the level of molecular biology as well as mechanisms for heredity and transmission of information from the lowest to the highest levels of organisation in an organism. I do not wish to discuss here the scientific import of this latter ideal.[188] What interests me here is the extent to which (1) these common goals help transform a case of disunified big science into a highly collaborative project, at least when compared to other examples of big science in biology and (2) such extensive and centralised collaboration impacts the type and quality of biological understanding obtained by the researchers involved.

Providing answers to these two key questions is, of course, very difficult, especially in view of the little empirical data yet available on the sociology of the Arabidopsis community. As already specified in Chapter 4, my own research on the collaboration patterns characterising Arabidopsis research has been focused on a very restricted set of powerful players in the community. Sociologist James Evans is the only other researcher, to my knowledge, who tried to study – and, in fact, measure – collaboration patterns among Arabidopsis researchers, specifically in relation to their ties to the industrial world (Evans 2004; more on this in the following sections). However, there are no data as yet documenting the quality of relations between specific sub-disciplines within the whole community and the extent to which they adhere to the common 'share and survive' ethos. My research on this point is therefore based on in-depth interviews conducted with a small sample of Arabidopsis researchers, as well as on relevant archival and published documentation produced in the community (and the published results of similar in-depth interviews carried out on a sample of American researchers by Evans). Both interviews and archival documents confirmed that researchers strongly value the peculiarly collaborative ethos characterising the Arabidopsis community and believe that it played an important role in their research practices and relations to other research groups. In other words, Arabidopsis researchers see themselves as strongly epistemically dependent on the rest of the community, since they value collaboration as the best strategy towards obtaining significant results in their work.

The research area dealing with cold acclimation constitutes a telling example of the difficulties encountered when collaboration is constrained by geographical and national limits, rather than being promoted across the whole Arabidopsis community. Cold acclimation is a very exciting field in plant biology: it brings together microbiology, plant ecology and evolutionary biology in order to understand the mechanisms through which a

[188] A consequence of this rhetoric is visible in the gene-centric perspective uncritically adopted, as we have seen in Chapter 5, by The Arabidopsis Information Resource.

plant can adapt to the temperature of the environment. Starting from the observation that plants can adapt to resist drops in temperature (for instance, by increasing levels of sucrose and the flexibility of cellular membranes), researchers have looked for correlations between such metabolic and physiological 'cold responses' and changes in the gene expressions patterns of the plant. They identified a regulon (that is, a gene cluster responsible for the transcriptional activation of a specific function) that is responsible for imparting freezing intolerance, thus activating cold response pathways that change the metabolism of the whole plant (Thomashow 2001). The study of this regulon, known as CBF, is an example of an extremely successful collaboration among Arabidopsis researchers with different expertises and coming from different epistemic cultures. Yet, collaboration patterns did not develop as ideally as Arabidopsis researchers would have liked: in fact, in the late 1990s two different research groups were formed, one based in the United States, the other in Japan. Both of them ended up discovering the CBF regulon through the same, 'classical' approach (that is, by mapping one gene to one enzyme and seeing what happens) and publishing results at more or less the same time, thus engendering a climate of competition and hostility between the two communities.[189] Collaboration within each of the two groups conformed to the norms established in the community, and their results certainly started a fruitful debate on how the process of cold response regulation is actually distributed among various gene clusters, which are now seen as interacting among themselves as well as in response to different physiological or climatic conditions. However, international communication between these groups and the rest of the community was not good enough for the two groups to realise that they were pursuing the same phenomenon with the same tools: at least, they did not realise it in time to avoid the subsequent conflict and, as a consequence, did not manage to establish cooperative ties with each other.

Examples such as this demonstrate that the extent and manner in which the 'share and survive' ethos is respected and enforced by different laboratories within the community varies depending on the specific geographic area where the laboratories are located, the personal ties between members of any specific laboratory and researchers in the rest of the community, the social and political context in which research is developed, as well as the specific field or aspect of Arabidopsis biology studied in any one lab. This does not diminish the importance of the ethos of accessibility as a widespread pattern in the Arabidopsis community. Yet, it highlights the fact that many groups participating to Arabidopsis research are at the same time participating in other epistemic cultures, depending on the specific combinations of expertises of their members. For instance, a laboratory working on the ecology of Arabidopsis variants participates in the Arabidopsis community as well as in the community of plant ecologists and, possibly, the field of ecology as a whole. The possibility – in fact, the need – for each local group to ally itself to different communities (and thus different epistemic cultures) has a great impact not only on how closely these groups respect the ethos of each of those communities, but also on the overall social dynamics and attitudes within the group. Trading between the 'share and survive' ethos proposed by the Arabidopsis community and the 'publish or perish' mentality plaguing much of the extant research in the life sciences proves a difficult task

[189] The development of cold acclimation research has been described to me by Sue Rhee (who was one of the principal investigators involved in the project) in an interview dated 8 August 2004.

for many of the individual researchers and research groups studying Arabidopsis. Arguably, the research areas whose results are likely to be significant in fields beyond plant biology (such as microbiology) are the ones that present the most problems: while their studies are conducted on Arabidopsis, the validity of their results has to be assessed by researchers working outside the Arabidopsis community.

There are no obvious mechanisms through which Arabidopsis principal investigators (PIs) can enforce adherence to such 'ethos of accessibility' among their colleagues and pupils, other than public acknowledgment of the fruitfulness of successful collaboration and recourse to social skills – of which I shall say more in section 7.2. Further, it is very difficult to determine in which instances scientists do follow the norm of collaboration and in which instances they instead decide not to. As I already anticipated, there are tensions between groups within the community and areas where competition is stiffer and sharing data more difficult than in others. I shall focus on such tensions in more detail in the following section. By looking at some of the problems encountered in enforcing the Arabidopsis ethos, I hope to provide a realistic picture of the complex, yet highly efficient, ways in which the Arabidopsis community remains centralised and largely united by common goals and institution, despite the diversity of groups and interests that populate it. I thus hope to provide some insight on the impact that the communal community ethos and centralised organisation have on the quality of the understanding of plant biology acquired by Arabidopsis researchers. As I intend to show, the ethos proposed by the founders of the community continues to have a significant influence on its success in recruiting new participants, attracting funding and, arguably, engendering new discoveries.

7.1.2 Struggles Within the Arabidopsis Community

> *We have put forth a goal of no less than complete understanding of the biology of an organism; the only way to achieve success is to work together with the realization that we are all wedded to the same goal*
> *The Multinational Arabidopsis Steering Committee, 2002*

As implied in my characterisation as centralised big science, a little group of key institutions and sponsors plays a significant role in directing and administering research production within the Arabidopsis community as well as between members of the community and the groups that are external to it. These institutions and sponsors are what I call 'central' actors in Arabidopsis research. They play a role akin to the one of benevolent tyrant, holding great power and the final word on every important decision and strategy to do with research in the community, yet at the same time absorbing information and insights (allowing him to take such decisions) from the various groups of researchers working in the community. I shall start my analysis of this situation and its implications towards the gathering of knowledge on Arabidopsis by focusing on the impact of institutionalisation on social dynamics *within* the community: that is, the institutional set-up devised, at least in principle, to serve the community ethos and the problems and tensions generated by the power of such institutions over local research groups. The next section will instead focus on so-called 'external dependencies', that is on the constraints and requirements imposed on Arabidopsis research by agencies that are not directly involved in carrying out such research, but nevertheless retain a high level of power over its modalities and outcomes: its sponsors (both in government and in industry).

As outlined in Chapter 3, the Arabidopsis community is officially centred on the MASC, or Multinational Arabidopsis Steering Committee. Members of this committee range from 15 to 20 depending on the year and serve for a total of four consecutive years. They are elected by representatives of the most prominent and active laboratories within each nation and are responsible for reporting on national initiatives and results in Arabidopsis research, while at the same time absorbing inputs from abroad as well as organising multinational projects and initiatives together with other national representatives. Meetings of the MASC, both informal and formal such as the ones happening at the yearly International Meeting for Arabidopsis Research, help to keep Arabidopsis researchers in touch with each other and to coordinate research directions across the community. The MASC thus tries to make sure that no two laboratories are focused on precisely the same phenomena (or at least not with the same tools, thus risking useless duplication of results as in the case of cold acclimation research) and that new applications for grants and resources obtained by specific groups in the community are employed to study yet unexplored aspects of Arabidopsis biology, thus progressively filling the gaps in the knowledge hitherto accumulated. This function is greatly appreciated by Arabidopsis scientists, who see centralisation, as embodied by the MASC, as a way to keep some coherence in Arabidopsis research, with a view to integrating all the knowledge of the plant into an integrated understanding of plant biology.

Closely associated to the MASC is a huge umbrella project, funded mostly by the NSF and the European Commission, called ‘Project 2010’. This project is the successor to the AGI sequencing project (1996-2000), and is meant to enforce collaboration among microbiologists and developmental biologists working in the community: the putative, yet extremely ambitious, goal of this collaboration is to be able to outline the functional role of each Arabidopsis gene by the year 2010 (hence the name of the project). This project certainly does not incorporate all laboratories engaged in Arabidopsis research. However, it is worth mentioning its existence, as the extensive network of collaborations and centralised organisation that it displays are representative of the type of projects encouraged within the community. Its ambitious goals are also indicative of the tendency, by Arabidopsis scientists, to think broadly about the significance of their research – a tendency that is much criticised by scientists worried about the representativeness of Arabidopsis with respect to other organisms, but which also signals the willingness to overcome fragmentation and competitiveness in biology and start building an integrated understanding of organisms.

Let me now focus again on the MASC and specifically on its members. The number of MASC members per country is more or less proportional to the percentage of Arabidopsis papers that are published in each country. In 2002, there was one member for Australia/New Zealand; two for Canada; two for China; one for the European Union as a whole, which is important to note since the European commission has been the major sponsor for Arabidopsis research in Europe already since the early 1990s; one for France; one for Italy; two for Germany; one for Japan; five from the U.K. (a predominance due both to the size of Arabidopsis research there and the importance of the U.K. as a platform for MASC discussions); and, finally, two for the United States. This latter number might seem disproportionate, since the U.S. hosts a good third of total Arabidopsis research as well as most of the resources and coordinating initiatives fostering it (such as the ABSC stock centre, the twin brother of NASC, and TAIR). However, it is important to note that most representatives for other countries at the MASC have either been trained in the States or have spent a considerable time visiting there. This is a first sign of a reality that we will encounter again and again in this chapter, and which is of capital importance to understanding the social dynamics surrounding Arabidopsis research: that is, the US have no rivals in the quality and quantity of Arabidopsis research that they produce as well as in the resources (financial, human and institutional[190]) made available to scientists. They are followed closely by the United Kingdom, where most of European Arabidopsis research, as well as the NASC, is based. In fact, the very birth of MASC is largely due to efforts by the USA and UK to institute an English-language, official forum to organise, control and promote advances in Arabidopsis research.

Table 7.1. Most Cited Institutions in Arabidopsis Research. The predominance of North-American research is indisputable in the face of these data. Compiled by Evans (2006).

Table A.3. *Arabidopsis* Publishing Activity: Most Cited Institutions, 1974-2002

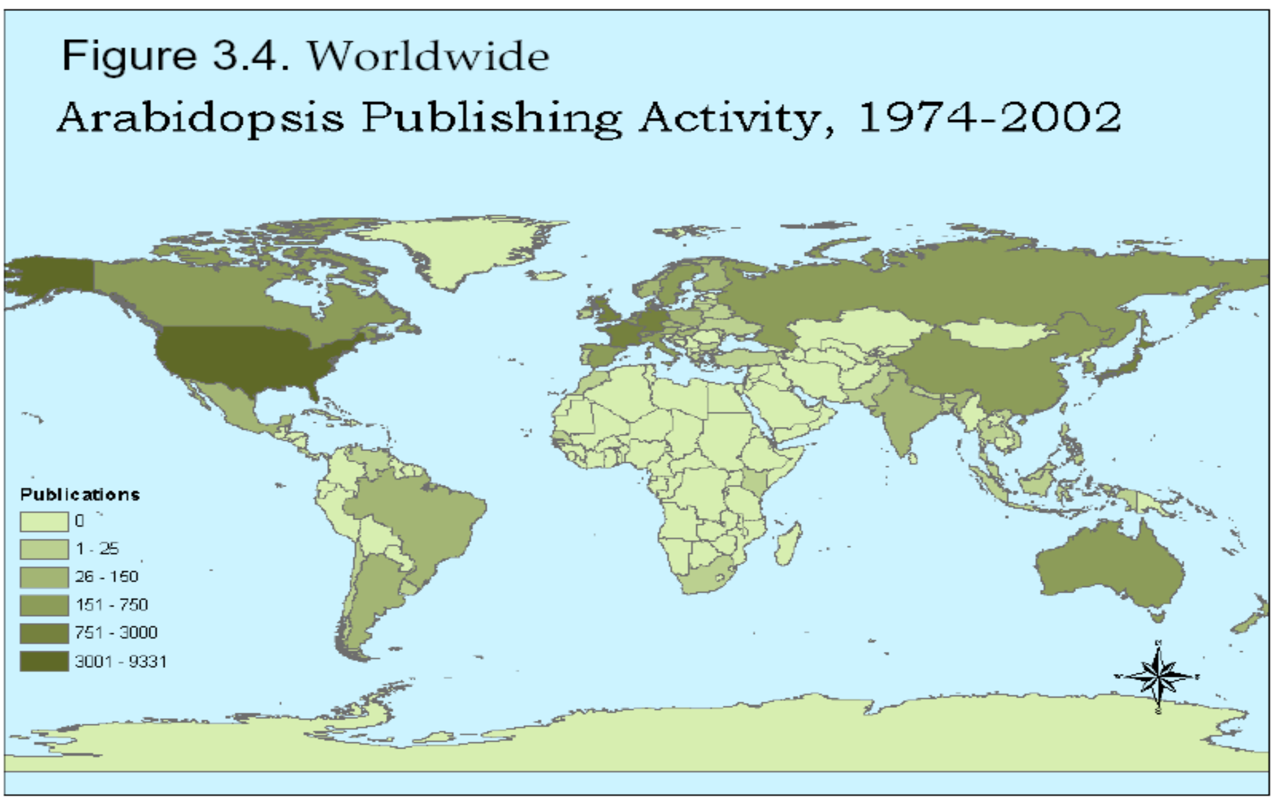
Institution	City	Most Cited Organization	*N* Cites [a]	*N* Papers [b]	Cites / Papers	First Paper
Michigan State University	East Lansing	Dept. Energy, Plant Res. Lab	14,707	510	28.84	1983
University of California, Berkeley	Berkeley	Plant & Microbial Biology	10,847	318	34.01	1989
University of California, San Diego	La Jolla	Cell & Dev. Biology Sections	10,407	269	38.69	1991
Caltech	Pasadena	Division of Biology	10,144	107	94.80	1984
Harvard University	Boston	School of Medicine	9,657	212	45.55	1988
University of Wisconsin-Madison	Madison	Department of Genetics	9,447	382	24.73	1987
State University of Ghent	Ghent, Belgium	Interuniversitair Inst. Biotechnologie	8,406	410	20.50	1979
INRA (Institute National de la Recherche Agronomique)	Versailles, France	Laboratoire de Biologie Cellulaire	8,371	396	21.14	1989
Massachusetts General Hospital	Boston	Department of Molecular Biology	8,079	142	56.89	1987
John Innes Centre for Plant Science Research	Norwich, England	Division of Cell and Dev. Biology	8,032	299	26.86	1991
University of Pennsylvania	Philadelphia	Department of Biology	7,051	129	54.66	1976
Salk Institute for Biological Studies	La Jolla	Plant Biology Laboratory	6,510	181	35.97	1989
University of Arizona	Tucson	Department of Plant Science	6,312	257	24.56	1988
Agricultural University Wageningen	Wageningen, Netherlands	Department of Genetics	6,387	194	32.92	1980
Stanford University	Stanford	Dept. of Biological Science	6,108	100	61.08	1988
Rockefeller University	New York	Plant Molecular Biology Lab	5,150	136	37.87	1988
University of Minnesota	St. Paul	Department of Plant Biology	5,084	176	28.89	1987
Cornell University	Ithaca	Department of Plant Biology	5,077	233	21.79	1991
Yale University	New Haven	Dept. of Mol., Cell and Dev. Biology	4,928	195	25.27	1979
Agricultural University Wageningen	Wageningen, Netherlands	Department of Genetics	4,860	91	53.41	1980
University of California, Davis	Davis	Department of Vegetable Crops	4,606	248	18.57	1990
CNRS (Centre National de la Recherche Scientifique)	Castanet-Tolosan; Strasbourg; Gif sur Yvette, France	Institut des Sciences Vegetales	4,132	180	22.96	1985
University of Tokyo	Tokyo, Japan	Institute for Mol. and Cell. Biosci.	3,999	303	13.20	1974
CSIRO (Commonwealth Scientific and Industrial Research Organization)	Canberra, Australia	Division of Plant Industry	3,838	134	28.64	1976
Ohio State University	Columbus	Department of Plant Biology	3,763	182	20.68	1978
Max Planck Institute	Cologne, Germany	Institut Zuchtungsforsch (breeding)	3,760	187	20.11	1990

[a] *N* Cites refers to the total number of scientific citations to *Arabidopsis* articles published from this institution.
[b] *N* Papers refers to the total number of *Arabidopsis* papers published from this institution.

[190] Note that, in the States, private foundations devoted to the financing of basic scientific research (such as the Carnegie Institute, where TAIR is based) represent a significant addition to governmental institutions such as the NSF or NIH. See Gerson (1998) on this.

American and English cultures are therefore bound to have a great impact on the values and norms implemented within the community as a whole, as well as on its research goals. Arabidopsis research is international insofar as the laboratories where it is performed are located in several nations around the world, ranging from the ones represented in MASC to smaller groups in Mexico, Chile, Brazil and Russia (see figure 7.1 below). However, it is important to keep in mind that almost half of these laboratories are in fact based in the USA or UK; that this gives a central role to research performed in those nations, which keep attracting the best researchers and resources to conduct especially molecular biology; and that this is bound to create tensions between Anglo-American ‘central’ institutions, holding much of the power and money required by contemporary science, and ‘peripheral’ institutions located in the remaining nations. A good example of this latter issue comes from France, where Arabidopsis research is also thriving, but where ‘francophone’ researchers have a notorious tendency to refrain from participation in international (and, as they see it, far too ‘anglophone’) projects – with the result that French contributions to the Arabidopsis community are fewer than what this powerful country could afford.

Figure 7.1. Distribution of Arabidopsis research worldwide up to 2002. Courtesy of TAIR.



As mentioned in section 2.3.2, it is hard to assess whether the type of institutionalisation that I characterised as centralised big science provides more advantages or more disadvantages to biological research practice and results. The problems with such an institutional set-up intensify, naturally, as the centre around which research is organised has very local characteristics tied with a specific epistemic culture, which therefore imposes itself on the rest of the groups involved. This is actually very often the case in centralised big science, as it is arguably difficult to institute central organs that are not influenced by a specific perspective, approach, methodology proposed by some of the most powerful members of the community. Admirably, an effort has been made within the Arabidopsis community to give a voice to all nations involved in this big science project: in this sense, the MASC is quite efficient in its setting up of representatives reporting on each nations' researchers' results, needs and wishes. This effort has led to many advantages involved in participation in the community, by enhancing epistemic dependences among Arabidopsis researchers of different nations and epistemic cultures. It also, as we shall see in the next section, improved the community's capacity to attract sponsors. Nevertheless, centralised big science can be very restrictive in the type of methods and results that it sanctions as acceptable. In the case of the Arabidopsis community, this is due to a variety of factors, some of which I shall now list and examine.

Let me start by observing that research in the Arabidopsis community typically involves several types of *peer-different collaboration*.[191] Perhaps the most common is collaboration among researchers with different disciplinary backgrounds, such as in the case of biochemists and developmental biologists working together to study Arabidopsis metabolism, or the case of mathematicians and ecologists exploiting each others' expertise to analyse variations among Arabidopsis mutant populations. We could define this type of peer-different collaboration as involving the coordination, and sometimes even the exchange, of different theoretical skills (such as the ability to juggle with mathematical symbols and to plot statistical studies based on empirical data about mutant populations). A second important type of peer-different collaboration consists of attempts to coordinate different performative skills in order to achieve a common goal. These collaborations bring together researchers with differing technical expertise (or, to use my own terminology, embodied knowledge). These include biologists able to analyse the same phenomena through different instruments, an occurrence that becomes more and more usual as instruments and techniques for both experimentation and data-analysis become more sophisticated, thus requiring unique and well-trained performative skills.[192] They also include collaborations between bioinformaticians (including programmers and engineers) and biologists, such as the ones I mentioned when analysing the TAIR project, which I will discuss in more detail below.

[191] See Chapter 2, section 2.3.1 for a discussion of Thagard's notion of peer-different collaboration.
[192] Consider for instance the different methods available to study gene functions: rather than training one individual in each Arabidopsis laboratory to use them all, which is anyhow unlikely, it is more profitable to institute collaborations among individuals endowed with different skills that are relevant to the research questions at hand.

The possibility to establish and carry out such collaborations is greatly enhanced by the central organisation of Arabidopsis research, thanks to which it is very easy to locate individuals with expertises and interests that complement one’s own in the required manner. Yet, peer-different collaboration encompassing the exercise of theoretical and performative skills involves a major difficulty, which Thagard aptly defines as follows: 'peer-different collaborators are exceptionally epistemically dependent on their co-workers, since they typically lack the skill to validate work done in a different field' (1997, 254). The issue of finding *common evaluative criteria*, which is indeed characteristic of interdisciplinary research as a whole, represents one of the most serious problems in the Arabidopsis community: as in the case of any other participant in big science, its scientific results and plans for the future need to be assessed over and over again in order to be sanctioned as appropriate and worth pursuing further. Up to now, it has been difficult to propose general solutions for this problem; as we have seen, that would imply agreeing on common standards, theoretical interpretations and research priorities, thus effectively dismissing the value of the diversity of approaches and insights characterising the community. The problem of finding evaluative criteria is therefore solved on a case-by-case basis. Often, however, such evaluations constitute strategic responses to extant social and economic issues at least as much as they address the scientific quality of research. As often pointed out in critiques of the peer review and funding allocation systems, the evaluation of the quality of projects might well turn into a power struggle among different parties, each of which represents its own interests and reasons for insisting on the adoption of its own evaluative criteria. The struggle is intensified by the fact that actors involved occupy a variety of positions in the strict hierarchy characterising academic science, some of which are immensely more powerful than others. PhD students and post-doctoral fellows, for instance, have very little decisional power over how their research will be interpreted and integrated into their supervisor’s project – no matter how great their contribution to the development of the research in question.

Table 7.2 – Academic Position of Subscribers to TAIR as for 2002 and 2003. Unfortunately many subscribers did not choose to specify their precise position in the academic hierarchy; however, these are the best data available to me at this time. Note that the number of senior and graduate researchers has been steadily growing from one year to the next. TAIR officials confirm steady growth also for years 2004 and 2005. Courtesy of TAIR.

October 2003		October 2002		
total community	12535	total community	11335	
Unspecified	8921	Unspecified	9438	
Graduate Student	737	Post-Doctoral Researcher	408	
Post-Doctoral Researcher	707	Graduate Student	372	
Professor	325	Assistant Professor	185	
Research Scientist	323	Research Scientist	163	
Assistant Professor	304	Professor	147	
Associate Professor	239	Associate Professor	135	
Research Associate	145	Group Leader	74	
Group Leader	125	Research Associate	74	
Research Assistant	111	Other	56	
Other	108	Research Assistant	48	
Unknown	97	Project Leader	43	
Research Fellow	70	Undergraduate Student	40	
Undergraduate Student	67	Research Fellow	32	
Project Leader	66	Director	28	
Director	41	Lecturer	23	
Lecturer	35	Curator	15	
Senior Research Officer	25	Unknown	15	
Curator	20	Senior Research Officer	14	
Programmer	16	Programmer	10	
Coordinator	9	Senior Lecturer	4	
High School Teacher	9	Teacher	3	
Senior Lecturer	9	Advisory Board Member	2	
High School Student	8	Coordinator	2	
Teacher	8	High School Student	2	
President	7	High School Teacher	1	
Advisory Board Member	2	President	1	
Middle School Teacher	1			

In the absence of detailed empirical data concerning the whole Arabidopsis community, let me exemplify these concerns through the cases of the research groups working at TAIR and NASC. TAIR lab members work together as a tight unit. Knowledge and tasks required for the creation of databases are necessarily distributed, as no single member of the team is knowledgeable about all disciplines involved: this creates a need for close collaboration and frequent meetings and informal exchanges within the group. This is evident in the decision, taken in 2004, to bring all collaborators of TAIR to the Carnegie Institute of Plant Biology in Stanford – a significant change from the original set-up, according to which half of the group (especially programmers) was based at the Santa Fe Institute in New Mexico. Work used to be organised via weekly meetings of the team at the Carnegie, in which the content of the tasks to be performed was decided upon. The two groups would have joint meetings, via conference call, every fortnight. This arrangement was discovered to be highly inefficient and needed to be changed. As also argued in programmatic statements featured in renewal NSF grant documents, one thing learnt by the first five years of TAIR experience is that the physical presence of all participants in the same institute is crucial. This finding is significant for Arabidopsis research as a whole, as it signals the great difficulties encountered when carrying out research collaborations among individuals located in different regions, or even different countries or continents. The optimisation of expertises involved in any particular project often goes at the expense of the optimisation of working conditions (geographical proximity, access to common facilities and so forth). A trade-off between these two important factors is thus a common characteristic of peer-different collaborations in the Arabidopsis community.

Another interesting feature of collaborative work at TAIR is a constant tension in the expertise and roles of what is regarded as 'technical' staff (that is, researchers whose theoretical and performative skills lie in the development of software and bioinformatics) and 'research' staff (that is, researchers whose theoretical and performative skills lie in molecular and developmental biology). The tension between the two groups is already evident from the classification of bioinformaticians as 'technical aid', rather than as researchers in their own right. This is in part justified by the guiding role assumed by biologists in developing TAIR. They are representatives for the prospective users and know what needs the resource is supposed to serve: thus, they are the ones developing visions for the users and evaluating the extent to which bioinformaticians can match those visions. However, as observed in Chapter 5, 'technicians' at TAIR are anything but passive instruments for the realisation of biologists' goals. They have their own expertise and goals, not to mention career prospects, given that bioinformatics is quickly becoming one of the hottest areas of research in computer engineering. They also strive to learn the biology involved in TAIR images and provide their own perspective on the alternative ways in which such knowledge could be visualised. This type of peer-different collaboration is, however, made difficult by the differing status and recognition reserved to biologists and informaticians within TAIR.

Last but not least, consider the extent to which not only internal, but also external collaborations sustains TAIR work. The decision-making process through which schemas, nomenclatures and data to include in the databases are chosen relies on:

(i) specialised experimental knowledge from published literature as well as consultations with experts in each field;
(ii) personal experience and affiliation of each TAIR curator with specific fields of biological research;
(iii) research conducted within the GO community, notably based in Cambridge (UK), whose researchers are actively collaborating with TAIR researchers (joint, cross-continental meetings are held every two to four months);
(iv) innovations produced by the bioinformatics community; and
(v) research conducted jointly by all curators of model organism databases: personnel from each prominent model organism community is in fact involved in frequent meetings in order to compare their results and strategies, as well as the prospects to develop a joint database collecting data from all organisms (to be called Generic Model Organism Database, or GMOD).

In the context of this extensive network of collaborations within and outside the Arabidopsis community, it is even more surprising to discover that the work of TAIR curators themselves is actually not valued by other Arabidopsis researchers, who see this kind of non-specialised research as a service to the community and, thus, as not leading to concrete results in specific areas of inquiry. In other words, TAIR curators are not seen as biologists and their work tends to be stigmatised, rather than applauded, by other researchers. This makes TAIR curators' position within the community rather peculiar, as their commitment to the integrative ideals underlying the TAIR project is counter-balanced by the certainty that, once that temporary project is terminated, it will be very difficult for them to find a good job within the Arabidopsis community of hands-on researchers.

Similar issues can be observed in the NASC laboratory, where we find three distinct expertises, whose boundaries are neatly demarcated: the biologists and technicians taking care of plant specimens; the biologists and bioinformaticians working on NASC databases such as PATO; and the technicians developing instruments for use by Arabidopsis researchers. Sean May, the director of NASC, explicitly told me that he searches for employees who are 'intellectually unambitious', that is, who can deliver what they are asked for in the simplest possible manner and who are uninterested in an academic career. This is because NASC work, requiring so much collaboration among biologists, bioinformaticians and technicians, can be frustrating to what he calls 'academic' mentalities: in the eyes of many Arabidopsis scientists, the goal for NASC is not so much to develop new knowledge, but to preserve and improve the conditions under which knowledge is produced, thus spending much time on elaborating instruments, tools and protocols.[193] Of course, what is also true is that May reserves for himself the privilege to supervise and assess research conducted in all three areas covered by his laboratory. As he told me, he has the perfect background for this job: a degree and interest in population genetics, expertise with transactivation techniques and a hobby in computing. He practically functions as a mediating figure between scientific and technological concerns in the lab. May's attitude stands for a different way to cope with the difficulty of evaluating and coordinating interdisciplinary research – that is, reference to one or more supervisors whose expertise covers all of the various fields involved.

The latter situation immediately engenders a worry: by assigning evaluative powers to just a handful of individuals, don't we risk to distort the results of Arabidopsis research to fit their individual beliefs and perspective? Indeed, as we saw in the case of TAIR, the risk of relying on a centralising theoretical perspective is not only present, but actually unavoidable, as reliance on a specific viewpoint is needed as a platform for the development of an integrated view on Arabidopsis biology. I shall expand on the epistemological implications of this claim in Chapter 8. What I would like to examine now is the extent to which one specific view, that is the gene-centric approach, continues to hold sway among the most prominent members of the Arabidopsis community. To start with, genome sequencing is generally presented as the necessary 'raw material' (i.e. as a starting point) in order to understand developmental processes. This is a highly reductionist assumption, as DNA sequences are but one component of an extremely complex ensemble of processes: it is the interaction of DNA with other cellular components, rather than its isolation into seemingly autonomous genes, that should be seen as the 'raw material' for the study of plant development. Furthermore, microscopic and mechanistic explanations of developmental processes are vastly predominant and sought after.

[193] Of course, it could be argued that the production of adequate research tools is as important to scientific output as experimental research done through those tools. Yet, a distinction between these two complementary types of activity seems to persist both within and outside of biology, at least in terms of the different levels of social status and prestige associated to each of them (Radder 1996, 43-44).

The emphasis on making sense of genetic sequences does not diminish the importance of the blossoming side of research focusing on higher-level organisational processes (e.g. embryological and growth processes). TAIR director Sue Rhee herself, despite co-ordinating what is still largely a collection of databases concerning the genomics of Arabidopsis, is a developmental biologist and is pushing the TAIR to integrate data from higher-level studies, from cellular to environmental. However, the pursuit of gene-centric research is today the most likely path towards earning resources and funding in Arabidopsis research. This is where governmental and industrial sponsors most like to invest; and in fact, it is this area that exhibits the highest condensation of charismatic and capable principal investigators, exercising a great degree of authority over their own laboratories, over broad research trends within the community as well as over funding agencies. The swaying authority of gene-centrism in biology is indeed, as remarked by many an historian[194], due as much to the scientific characteristics of this view (including its relative simplicity) as it is due to the central role played by its spokespersons within the biological community at large. This is evident also in the Arabidopsis community, where gene-centrism and powerful, charismatic scientists seem to go hand-in-hand, not least because groups that do not follow the gene-centric trend find it still difficult to acquire funding, resources and visibility in the community.

Power struggles among different groups in the Arabidopsis community are often not based around competition for results, resources and recruits, but rather around the quest for authority among one's peers. Among others, sociologist Pierre Bourdieu highlights this point in his definition of a scientific field:

> As a system of objective relations between positions already won (in previous struggles), the scientific field is the locus of a competitive struggle, in which the specific issue at stake is the monopoly of scientific authority, defined inseparably as technical capacity and social power, or, to put it another way, the monopoly of scientific competence, in the sense of a particular agent's socially recognised capacity to speak and act legitimately (i.e. in an authorised and authoritative way in scientific matters) (1975, 19).[195]

The gist of Bourdieu's argument here is that whenever a scientific research group forms around the study of specific phenomena with a common set of theoretical and methodological tools, there is chance for competition among its members over whose approach and insight will direct the group's research direction in the years to come. Figures like Chris Somerville provide a good illustration of the gains deriving from winning the battle over authority in the field. The importance of retaining influence over others in the same community becomes even more evident when considering competitions for authority among different groups, rather than individuals, within the community – which brings us back to the issue of centre versus periphery of research from which I started my analysis in this section.

---

[194] Keller (2000) and Oyama (1982).
[195] See also Bechtel 1993, 280 and Latour and Woolgar's discussion on credibility (1979, ch.5).

It would in fact be wrong to characterise Chris Somerville as a lone, gifted scientific hero struggling for recognition in the face of hostile academic circumstances. Somerville had many advantages to start with over many of his colleagues. He was trained in the best schools of what is currently the most powerful (and wealthy) nation on Earth and, when still very young, he was assigned the direction of a whole laboratory and was given relatively large freedom to pursue whichever research direction he chose. These advantages were not undeserved: he had proved from a young age to be an exceptionally talented and dedicated individual, whose many gifts extended to personable, friendly manners and 'a way with people' (Koornneef, pers. comm.) that gained him many a powerful friend in governmental circles. He was clearly the right man for the job and made the most out of the resources that were given to him. Yet, the fact remains that those resources were available to him, as was the possibility to communicate with English-speaking colleagues in his own language, form friendships that would greatly enhance his scientific authority and work in institutions with a reputation for scientific excellence.[196]

Chris Somerville's group is in fact one of the most influential in the Arabidopsis community as a whole. Members of his laboratory include some of the best Arabidopsis researchers in the world, several of which have been or are now acting as elected members of MASC, the steering committee coordinating Arabidopsis research worldwide. TAIR itself is headed by a student of Somerville's and TAIR activities are conducted in the same building as his research laboratory, so that there is constant feedback flowing from his office to the TAIR section (and from there, through TAIR databases, to the whole community). In short, Chris Somerville's Stanford group, including TAIR, constitutes what we call 'central' science in the Arabidopsis community. This is opposed to 'peripheral' research, that is research conducted in areas and institutions that are not as recognised and valued – in a word, not as authoritative – as Stanford University is in the academic world. These places are too poorly equipped to guarantee cutting-edge results (most of which require access to the latest technology); they have little funding to hire new talents and even less to sponsor their member's travels to the venues and meetings set up at more central places, for instance, by the MASC.

As an example of the issues about 'who is central' in the Arabidopsis community, and will thus influence research directions and results obtained from the study of Arabidopsis, consider the conflicts arising between the different databases organising Arabidopsis data as a service to the community. As we saw, TAIR is indisputably the best organised and most 'central' of these projects. However, this does not mean that TAIR is also the most used database for Arabidopsis data. In fact, researchers interested in more detailed data on Arabidopsis genomics might prefer to resort to the MIPS; researchers interested in the morphological characteristics of Arabidopsis plants, as well as their physiology, might instead turn to PATO, the database for plant traits and attributes developed by the NASC. Notably, MIPS makes little use of bio-ontologies and organises its data more loosely, yet

---

[196] Evans (2006) provides a very interesting picture of Chris Somerville as an entrepreneur, an activity which he successfully managed to couple with his academic position and which brings out many of his talents in dealing authoritatively with people.

in a way that is less dependent on the theoretical framework endorsed in TAIR. Per se, this feature does not make it incompatible with the TAIR effort, as demonstrated by the fact that many Arabidopsis biologists make use of all three databases in their research. At least in principle, all database projects (especially TAIR and NASC, because of their collaboration on seed distribution) are supposed to collaborate at all times, thus providing biologists with the best range of tools for the retrieval and digital handling of Arabidopsis-related data. Such direct collaboration does not, however, take place. Even if collaborating on the surface, there is hostility between the TAIR and the NASC group, which was visible in the interviews I carried out with the two directors of the projects, Sue Rhee and Sean May. Rhee did not even acknowledge NASC efforts at database development as a significant addition to TAIR's own efforts. May, par contre, was fairly critical of TAIR attempts to unify data under the same terminological banners, as he found that the price to pay for such integration (that is, the loss of pluralism of perspectives and of a great amount of partially, yet not entirely, overlapping data) is too high. Rhee also displayed a general disinterest for what researchers at MIPS were up to. She sees the MIPS as an offspring of the AGI that did not integrate very well into TAIR (she defined it as an 'alternative approach to TIGR', which is the database on which TAIR has modeled its efforts from its inception) and displayed reluctance in approaching the issue – due partially to her ignorance of what MIPS researchers are actually doing. Mostly, she was critical of the decentralised strategy adopted by MIPS, where collection of data is not ordered along pre-established lines, but is categorised progressively depending on which types of data become available.

The tension among strategies adopted by different Arabidopsis databases in the USA and Europe results in the different status of these databases within the two continents. While in the States the authority of TAIR seems undisputed, European Arabidopsis researchers are more aware of the existence and differences among the three resources and use them depending on their needs (it would take months of further empirical work to substantiate this claim, but this is the trend I saw emerging from my interviews). Already from this example we can see that the struggle for who is central in the Arabidopsis community is ongoing and has severe repercussions on the research being conducted. Ongoing disputes between central and peripheral loci of Arabidopsis research, as well as between different expertises represented within the community, indicate that the Arabidopsis community is, for all sorts of scientific as well as socio-economic reasons, very diverse and unequal in its access to resources, expertise, epistemic authority and research instruments. In other words, epistemological diversity depends not only on differences in skills and theoretical perspective: it also depends on the resources and social contexts available to individual researchers. In the midst of all these tensions, however, another important factor contributes to the enforcement of a common ethos and common goals throughout the community, thus maintaining its basic structure as centralised big science: that is, the nature and interests of its sponsors.

7.1.3 Heterogeneous Alliances: Governmental and Industrial Sponsors

A prominent factor allowing for the centralised organisation of Arabidopsis research has been the heavy involvement of governmental agencies as its sponsors. This is mainly due to four reasons. First, the government granted funding for basic, rather than applied, research, thus allowing Arabidopsis researchers to develop plans best suited to understanding plant biology for its own sake, rather than trying to acquire knowledge that could be exploited in the short term for agronomical or pharmaceutical purposes (as more commercially-oriented sponsors would have required). Second, governmental funding allowed for a centralised organisation of national resources on Arabidopsis research, thus preventing the fragmentation of results and projects that would have occurred if many different agencies had been involved. Third, research groups sponsored by the government found it much easier to implement the ethos of accessibility and data-exchange that would become a trademark of the Arabidopsis community. This would have almost certainly been impossible in the profit-oriented, competitive environment characterising privately funded research, as researchers working for firms are usually required to agree that disclosure and/or publication of results be agreed upon by the firm, on a case-to-case basis. Last but not least, the reliance on governmental funding allowed Arabidopsis researchers to transform requests for grants into opportunities for international alliances, both at the scientific and at the political level. As both Chris Somerville and Koornneef told me, funding applications in the USA would always point to the need to match the resources employed in the UK on Arabidopsis research. Researchers would emphasise how failure to obtain governmental funding would imply being left behind by European scientists: not only would US scientists be unable to compete, but they would also lose the opportunity to collaborate with other countries (as they would have little to offer to such collaboration). This strategy worked as efficiently in the US as it did in Europe and the rest of the world, where governments were pressured into giving money to keep up their scientific reputations. Allocating suitable funding for Arabidopsis research became a matter of national pride and strategic international collaboration, rather than a purely scientific issue.

But how did this caricature of an arms' race begin? The circumstances of initial NSF involvement in Arabidopsis research are relatively clear. The main force behind this event was James Watson, discoverer of the double helix and, by the early 1980s, one of the most powerful men in American science. Convinced that Arabidopsis could indeed constitute an extremely fruitful model for plant biology and microbiology at large, Watson convened a meeting in 1989 involving several representatives of the NSF, among whom director Eric Bloch, and several prominent Arabidopsis scientists, among whom Chris Somerville and Elliott Meyerowitz. Following this meeting, NSF officials offered large support to the ambitious proposals presented by the scientists: in fact, it was then that projects such as the construction of a centralised stock centre, databases and coordinating institutions such as the MASC were presented and initiated. This strong impulse matched the existing programs already implemented by the British government and the European Commission. Together, these funding programmes constituted the start of an international effort to support Arabidopsis research, which culminated in the successful collaboration characterising the sequencing project.

One of the effects of the alliance between prominent Arabidopsis scientists and governmental agencies is the concentration of key roles within the community in the hands of few researchers. Take again the case of Chris Somerville and Meyerowitz. Besides holding key academic positions and conducting research in important fields (Meyerowitz on flowering and Somerville on cell wall and plant body development), Somerville and Meyerowitz are key members of the scientific committee advising the National Plant Genome Initiative, as well as privileged advisors to the NSF funding committee. They also play a key role as spokespersons for two of the major web resources on Arabidopsis: Somerville is the scientific patron of TAIR (whose curator Sue Rhee is a former student of his) and Meyerowitz is the founder and current editor of The Arabidopsis Book.[197] Chris Somerville is acting as consultant for a variety of private funding agencies, such as, most remarkably, the Bill Gates Foundation.[198] This latter responsibility seems to confirm the full realisation of the vision that Chris and Shauna Somerville had of Arabidopsis research at the very beginning of their career: they are now in a position to act as 'expert advisors' for projects that concern applied Arabidopsis biology, most notably biotechnology and the production of GMOs.

In fact, it is now time to add some layers of complexity to the simple picture of alliance between governments and science that I painted here. Complications consist chiefly in the increasing interest expressed in Arabidopsis by commercially-oriented firms, several of whom understood already in the 1980s that research on this plant would have earned academic science scores of important inventions in molecular biology – inventions whose patenting would have earned universities the right to exploit their commercial applications. Biotechnology corporations such as Monsanto, Novartis, DuPont and Dow thus started to compete with academic research in order to develop methods to genetically modify plants for agronomic purposes.[199] This research was secondary in size to the Arabidopsis research carried out through governmental sponsors, both because only few companies decided to hire their own researchers to carry it out and because the industries that offered to sponsor university programmes had to sign agreements favouring disclosure (thus promising to distribute research results to the rest of the community, in conformity to the Arabidopsis ethos).[200] However, commercially sponsored research contributed in significant ways to the study of Arabidopsis. Perhaps the most significant example of such a contribution is the development of a key technique for the mutation of Arabidopsis plants, that is, the infection of plant tissues with Agrobacterium tumafaciens (a discovery that, as we saw in Chapter 3, had a tremendous significance for the expansion of Arabidopsis research): the technique was first proposed

[197] For information about TAB, see Chapter 4.
[198] Chris Somerville confirmed this engagement to me during an interview held on the 23rd of August 2004.
[199] Evans reports a quote from the chief technology officer of one plant genomics company founded in the early 1990s: 'being a company... you've got to worry about the competition and the competition is really the public sector' (Evans 2004, 61).
[200] Evans draws a distinction between two modes of interaction between corporate sponsors and scientific researchers: *relational ties*, signalling 'joint company-university research' and *transactional ties*, indicating 'exchanges of company money for academic results' (Evans 2004, 8; 2006a). The number of subgroups within the community engaged in relational ties is, for now, relatively low compared to the research areas sponsored by, and carried out within, academic institutions.

by a Monsanto laboratory, then expanded by Ken Feldmann and David Marks under the Sandoz Crop Protection Corporation. Another contribution was the donation of data acquired through corporate sequencing projects, whose results were used to refine the sequencing data already obtained by universities through the AGI project. Last but not least, industrial research has successfully targeted the need for adequate research instruments by academic scientists. The best example of this is the development of chips serving microarray research, which now go under the name of Affimetrix, or 'Affy-chips', from the name of the company that invented and distributes them (Evans 2004, 63).

What impact does the involvement of industrial interests have on the conditions through which Arabidopsis biology is investigated and understood? This question has no simple answer. It is certainly true that most strands of basic research in Arabidopsis are intermingled with considerations about applicability to biotechnology, or even with applied research itself. Scientists seem not only to be aware, but also to be proud of the overlap of basic and applied research, both because they seem to believe in its intrinsically positive ethical character and because it boasts the funding that governmental and private agencies are willing to provide to biological research.[201] Note that Arabidopsis research is a major part of the US-led political and scientific effort towards the enhancement of agricultural productivity, an effort that includes the study of 'optimal' cultivation techniques and the use of genetically modified organisms. The rhetoric of benefiting mankind by enhancing plant productivity and defeating plant diseases is used uncontroversially in most major scientific writings (and grant applications) on Arabidopsis. Consider the following extract from an important review paper published by Somerville in 'Science': 'Because as much as 40% of plant productivity is lost to pests and pathogens in Africa and Asia, strategies to control these will be a continuing challenge. The recent progress made in understanding the molecular basis of race specific pathogen interactions provides a new basis on which to try and develop durable forms of resistance' (Somerville 2000, 22).

The ethical questions arising from this situation are major ones, given the widespread opposition to the policies of dissemination of such agricultural practices in both Western countries and the so-called Third World. The production of GMOs is unfortunately coupled in most political domains to the old-fashioned discourse of 'development', defined as the duty by Western countries to help 'less developed' (i.e., culturally and economically different) countries by exporting and enforcing their own technology (on the acknowledged failure of such discourse of development, see Rich 1994). Technological and scientific knowledge are routinely presented as coupled with the values of representative democracy and the neo-liberal ideals of efficiency and production (Escobar, 1994). Both academics and activists all over the world have defended the value of non-western agricultural practices on the social ground that they are intrinsically valuable as part of the local cultures, as well as on the scientific ground of pointing to the value of natural diversity versus artificially manipulated diversity (see Radder 1996, ch.7; Shiva and Holla in van der Zwaan and Petersen 2003; Shiva 1999,

[201] This viewpoint is common to all American biologists that I have interviewed.

2002). It is in fact the case that most GM cultivations have hitherto been employed as monocultures, to be used in large plantations (thus favouring corporate over local administration) and to take the place of modes of production enhancing diversity and rotation of cultivations (e.g. soy cultivations in North Africa, China and Philippines). Further, the rights towards the production and distribution of GM seeds fell in the hands of only five major corporations, each of which already infamous for its lack of environmental ethics.

Little is known about the effects that such controversies on the ethical value of Arabidopsis biology and its applications have on the practices and results of Arabidopsis researchers. Certainly, the public image of plant biology has been affected. This is especially true in Europe, where a moratorium on GMO research issued in 1999 has caused difficulties even for basic research on plant biology (Helleman, 2003). Unfortunately, without appropriate empirical research it is impossible to determine the extent to which Arabidopsis research is materially constrained by these controversies on its commercial significance and humanitarian aspirations. The same holds for personal views and involvement of Arabidopsis scientists. The extent to which scientists are involved, personally or publicly, in controversies over the applications of their research remains an important open question when trying to assess which social parameters are relevant to the acquisition of knowledge about Arabidopsis.[202]

As mentioned above, James Evans has been carrying out substantial research on one related area, that is, the extent to which industrial ties and corporate sponsorship affect the content and distribution of Arabidopsis research. His results point to the following two trends:

(1) on average, affiliation with industry makes science ‘less novel and more commercial’ (Evans 2004, 6): it discourages scientists to repeatedly test their findings before publication and runs counter the ethos of accessibility pervading the academically-sponsored parts of Arabidopsis research (ibid.,198);
(2) while providing a wealth of additional resources for further research, private funding is usually highly unequal in its distribution – most of it is assigned to central loci of Arabidopsis research, to the obvious expense of the periphery (ibid., 6; as I discussed above, this can also be said of governmental research).

As put by Evans, ‘by expanding the web of science, industry involvement also stretches it thin and makes it fragile’ (2004, 6): social, economic and political considerations tend to override concerns about the quality of the research that is conducted as well as the reliability of the results that it yields.

The scarce presence of research that is directly controlled by industry is a main reason for the success and endurance obtained by the distinctively accessible ethos of the

[202] Another interesting question awaiting further research is whether, depending on variations in location, epistemic culture, interests and so forth, different groups within the Arabidopsis community have different perceptions of the ethical and commercial value of their research.

Arabidopsis community as a whole. However, it is important to note that industrial interest – and thus sponsorship – in Arabidopsis is increasing fast, due to the more and more obvious applicability of the results obtained by academic researchers. The academic community is also becoming more sensitive to the opportunities and research directions offered by corporate funding. This is partly due to the constant difficulties experienced by senior Arabidopsis researchers in renewing sponsorship from governmental agencies. Yet, it is also a signal of the increasing willingness by the government itself to sponsor applied over basic research, thus reaping the dividends of a quarter of a century long investment in basic plant biology.

.

7.2 The Significance of Social Skills

> *If someone is comfortable with the things and language used by a group of others, we say that he or she is a member of that group. In this sense, categories – our own and those of others – come from action and in turn from relationships*
>
> *Geoffrey C. Bowker and Susan Leigh Star 1999, 285*

On the basis of the observations made in the previous section, and following up on the considerations hitherto made on the nature of understanding, I would now like to argue that participation in one or more scientific communities, and thus communication and exchange with other scientists and affiliation with one or more epistemic cultures, are an essential condition for acquiring *scientific* understanding. This might not seem to be the case when looking at my analysis of theoretical and performative skills. After all, I hitherto qualified the coordination of theoretical and embodied knowledge (that is, the main condition I posed for the acquisition of understanding) as the result of individually performed skills. Further, my definition of understanding itself does not seem to leave much space for the social realm: if understanding is a cognitive process, how can it be shared with others and, most importantly, how can this matter to the quality of the understanding that is achieved? In this section I intend to answer these questions by pointing to a third set of skills whose exercise I take to be indispensable to defining understanding as ‘scientific’. These are what I call social skills, that is, the ability of an individual scientist to exchange knowledge, both theoretical and embodied, with other scientists – including the theoretical and performative skills that we have seen to play a crucial role in the understanding of phenomena.

There is a sense in which my analysis might seem trivial: after all, it is rather obvious that the pursuit and accumulation of knowledge requires interpersonal communication. Nevertheless, this important point often plays the role of ‘elephant in the room’ within philosophical discussions of scientific epistemology. It is not explicitly acknowledged and, most importantly, its consequences are not investigated. For instance, Miriam Solomon discusses the social nature of scientific practice by pointing to what she calls ‘non-empirical decision vectors’, i.e. factors that inform individual’s decisions concerning their research, even if they are independent from the empirical success of such research (Solomon 2001, 57). Examples for these vectors are epistemic values like simplicity and elegance, but also personality traits such as pride, conservativeness or radicalism. Through her framework, Solomon ends up depicting scientific practice as an ensemble of individual actions and decisions: she does not consider the manners in which such actions affect each other, nor does she highlight how constraints on communication (such as, most remarkably, institutional and material constraints) influence individual decisions.

In the analysis that follows, I intend to expose two crucial roles played by social skills in shaping the understanding of a phenomenon that is acquired by an individual scientist – and in fact, as I will argue, in transforming it into 'scientific' understanding. One is to enable access to the theoretical and performative skills needed to understand a biological phenomenon. The exercise of social skills is crucial to the successful accomplishment of activities such as conquering recognition and authority within a research community or field; acquiring resources; participating in appropriate networks; earning prestigious job appointments; and gaining access to laboratory materials – elements that, as we have seen in the previous section as well as in the last section of Chapter 6, are part and parcel of scientific research as a whole and particularly of research organised as centralised big science. The second important role played by social skills is to allow researchers to communicate their insights to others. Social skills enable a scientist to secure a material, social and economic environment that is suitable to creating tools (such as models, theories or instruments) that might help others to acquire and exercise the same theoretical and performative skills as the ones exercised by that scientist in the first place. Once other researchers are able to interact with phenomena under the same conditions and with the same skills as used to acquire a specific type of understanding, they will be more likely to undergo a similar experience of understanding.

In this section I will examine both of these roles in turn, first by analysing their general features and then by providing examples of social skills that fit one or both of these roles, as exercised in the context of Arabidopsis research. I shall start with the social skills relevant to the individual acquisition of understanding and then proceed to examine the social skills required to communicate such understanding to others. My list of social skills is by no means complete and its relevance to research practices in other communities remains to be investigated, given the extreme variability of social dynamics characterising research communities in different disciplines and fields. Further, most items in the list are closely related to each other, sometimes to the point of partially overlapping (as in the case of personability and charisma, for instance) – an occurrence due to the fairly limited and highly institutionalised set of activities, or 'research practices', within which they are exercised. I hope that this list can nevertheless serve as an illustration of the relevance of social skills to the acquisition of scientific understanding.

7.2.1 *Social Skills Towards Acquiring Understanding*

Social skills enable scientists to acquire theoretical and embodied knowledge of a phenomenon and thus to perform the right epistemic activities towards understanding it, including for instance modeling or experimental interventions. A scientist that is adequately socially skilled will be able to participate in the research communities that provide the theoretical and performative skills needed to investigate the phenomena of interest to him or her. As becomes evident from my analysis in section 7.1, participation in a scientific community (e.g. training in its adopted theoretical and embodied knowledge, including theories, methods, tools, ethos and experimental results) is crucial to being able to identify a phenomenon to be investigated as well as the means through which such investigation should be carried out. A researcher or layperson that is not exposed to the epistemic culture of the Arabidopsis community (and, at the local level, of some of its subgroups) would not even be able to decide which aspect of Arabidopsis biology to investigate, nor to use knowledge previously accumulated on the same topic as a starting point for his or her own research. The acquisition of adequate social skills is made even more relevant by the centralised nature and common ethos of the Arabidopsis community: for instance, not only it is important to participate in at least some of the community events organised on a monthly or yearly basis (such as the International Arabidopsis Meeting), but, once there, one needs to be able to express one's ideas and results with reference to the work carried out by others, as well as the values and motivations underlying such research.

Remarkably, social skills are useful to acquiring all kinds of theoretical and embodied knowledge, whether such knowledge is acknowledged and approved of within one or more scientific community or not. This is fine for my purposes, since I do not think that *all* the knowledge used in acquiring understanding must necessarily come from scientific training and exposure to scientific epistemic cultures. The whole experience of an individual may be relevant to the understanding of a phenomenon: it is hopeless to try and demarcate between ways of thinking and acting acquired through scientific training and ways of thinking and acting learnt in everyday life and informed by personality, talents and upbringing.[203] In fact, I would argue that social skills represent a *modification* of skills acquired in everyday life *with the aim to fit* the epistemic culture(s) characterising the scientific communities in which researchers find themselves. This is why many of the skills listed below are not by any means specific to science: rather, they are exercised in specific ways in different specific communities because adequate exercise of these skills is required for scientists to acquire the theoretical and performative skills needed to carry out their research. In fact, there is a sense in which social skills are subsidiary to theoretical and performative skills in science: they are not part of the knowledge that researchers aim to acquire, but they constitute a necessary means towards yielding scientific knowledge.

Planning

[203] The theoretical and performative skills learnt in scientific education are not always exclusive to science: they represent forms of everyday skills that are refined and adapted to the specific concerns and requirements of investigation of natural phenomena through specific (conceptual or material) tools and instruments. This is true, for instance, of the skills required to abstract. While these are specifically modified to help in constructing, handling and interpreting models, the ability to abstract features of the natural world to construct representations of it is by no means limited to science: rather, the use of the same basic ability (as for instance in the arts or in street signs) is modified to fit different concerns from the scientific ones.

Strategic thinking and planning is a fundamental skill for researchers to learn. The rational vision constructed by Chris and Shauna Somerville at the start of their career in Arabidopsis research constitutes a blatant example of the usefulness of planning adequately one's training and career moves: they were both successful in working within environments that welcomed their attitude and ideas, as well as providing space for experimenting not only with a novel model organism, but with a whole new approach to research in molecular biology. More generally, by rationally choosing which social and scientific environments to participate in and contribute to, biologists actually select the skills that they might want to learn so as to pursue their research goals. I shall come back to the significance of picking and developing specific skills when discussing the related commitments in Chapter 8.

Competitiveness

Biologists working in big science projects need to be able to withstand an often aggressively competitive research environment, in which scores of highly qualified individuals study the same phenomena and strive to produce results as fast as possible. Typically, this means that individuals who display some propensity towards competitiveness might acquire more resources, power and/or visibility than individuals who refrain from the spotlight. Within the Arabidopsis community the situation is made more complex by the predominance of the ethos of accessibility. As we have seen, such an ethos considerably lightens the competitive climate in Arabidopsis research is conducted. However, some areas remain competitive, either because of their own internal dynamics or to withstand pressure from the rest of the biological community. In such cases, in order to preserve appearances, attempts at competition have to be carried out without direct confrontations but, when possible, under the pretence of collaboration.

Open-Mindedness

Open-mindedness might seem an unlikely candidate for a social skill, as it concerns the ease with which researchers deal with new ideas, projects, methods and so forth. Yet, new ideas usually come from other people or tools for mass communication – colleagues, journals, conference talks, textbooks. An unwillingness to consider new and possibly different viewpoints could seriously damage the education of a researcher, barring him or her from considering alternative interpretations of a phenomenon or alternative strategies through which to study it. The very participation in Arabidopsis research was, for biologists joining at the beginning of the 1980s, a considerable gamble motivated by their ability to envisage the possible advantages and new opportunities offered by the new model organism (and, for animal biologists, for work on plants in general): the exercise of open-mindedness paid high dividends for people who took that decision. Also, we have seen in the previous section how peer-different collaboration now constitutes a main avenue towards developing integration among different areas of organismal biology: the construction of an integrative framework also requires, first and foremost, the ability to learn from colleagues working on one's same topic with differing tools, perspectives and expertises.

Patience and Dedication

The abilities to be patient and to persevere with the study of a single topic would seem, prima facie, to constitute good candidates for performative, rather than social, skills. In fact, I regard them as valuable in both senses. They help researchers to interact with their material environment in ways that are helpful to their gaining an understanding of a phenomenon, thus playing the role of performative skills (for instance, consider the patience displayed by a researcher in carefully sawing and growing Arabidopsis plants according to the protocols distributed by the NASC). Yet, they also enable researchers to absorb from their teachers, colleagues and students the theoretical and embodied knowledge that is necessary for them to perform their research, thus acting as a social skill. Training within scientific communities requires a great deal of patience and dedication, as does peer-difference collaboration (which includes a great amount of learning for all participating sides). Maarten Koornneef, now a central figure in Arabidopsis research, started his study of the plant at a time when few of his peers shared his conviction that plant molecular biology would eventually blossom. His patience and dedication over the first few years of his career in Wageningen paid off as soon as his results became known internationally, causing him and his lab to be hailed as a prominent source of data and expertise on Arabidopsis. Dedication is also visible in the case of Chris and Shauna Somerville, who devoted their whole life to uncovering the mysteries of plant biology.

7.2.2 Social Skills Towards Communicating Understanding

I have hitherto qualified understanding as a cognitive process that results from the skilful and efficient coordination of theoretical and embodied knowledge. Given its nature as a cognitive achievement, it thus seems strange to define such understanding as social in the sense of being communicable to others: rightly so, since there is no way, as far as I know, of ‘convincing’ someone to undergo the same cognitive process. Does this mean that the very idea of communicating understanding is unfeasible in my framework?

The direct interpersonal exchange of scientific understanding is in my view unrealisable, in the same sense in which other cognitive processes (such as seeing, listening, thinking) cannot be directly taught to or exchanged with others. However, there is a sense in which an individual can provide others with tools that will help them to experience the same cognitive process as themselves. Take the case of seeing. We cannot teach someone to see a hare hiding in the grass: either it happens or it does not. However, we can teach someone what to see and when by, for instance, pointing in the direction of the hare, drawing it on a piece of paper (so as to allow our interlocutor to identify it) or directing him or her towards the hare (even if that would probably cause the hare to jump away). In other words, we can instruct someone on the conditions by which he or she might see what we do. The same is true of understanding: we cannot make someone understand something in the way we do, but we can provide tools that will help that person to understand a specific phenomenon in a way similar to the way in which we have understood it. In science, this corresponds to trying to express and communicate the skills and commitments adopted by an individual scientist to understand a specific phenomenon.

Of course, this means that there can never be certainty as to how exactly any individual understands something, nor as to whether a group of people who professes to share the same understanding actually does. Yet, this is true of any cognitive process (and indeed, philosophers of mind have long referred to this phenomenon as the 'problem of other minds'). In the absence of a precise measurement tool, there seem to be only two guarantees of the quality of understanding acquired by an individual: one is the way in which her understanding of phenomena informs her interactions with others and with the material environment; the other, most interesting for my purposes, consists of reconstructing the conditions under which understanding has been acquired.

Let me now list some of the social skills that enable researchers to share understanding with one another.

Authority/Charisma

Again, the powerful protagonists of Arabidopsis research provide an excellent example of how the exercise of authority, united with the ability to exude charisma, might go a long way towards facilitating the development and implementation of research programmes, the acquisition of funding and the recruitment of pupils (not to mention enhancing the sympathy of peers and collaborators towards one's ideas and plans). As remarked by William Bechtel,

> There are both important social and cognitive components to the process by which scientists acquire the authoritative stature, which enables them to settle questions as what lines of investigations should be taken seriously and what techniques are appropriate. The social components include factors as the ability to interact with others in a manner that commands respect and to plot a course of action that will result in that respect. There are able practitioners of many research fields who lack such social capacities and consequently are handicapped in establishing authority. But there are also cognitive considerations in developing authority, such as successfully developing new techniques or respectable arguments for pursuing a line of research. (Bechtel 1993, 281)

Both Sue Rhee and Sean May, whose daily activities I had the chance to observe, exercise remarkable authority on the members of their laboratories, which is essential towards securing a harmonious and efficient working environment. They also see themselves as responsible towards the many scientists making use of TAIR (in the case of Rhee) and NASC (in the case of May). Following on Bechtel's argument, such a sense of responsibility is justified, given the great authority that they exercise (personally as well as through their scientific choices and claims) over the rest of the Arabidopsis community, particularly its periphery. They are responsible for developing tools, models and databases for use by any interested plant biologist: through this appointment, situated at the very institutional centre of the community, they influence the ways in which Arabidopsis researchers interact and perform experiments in fundamental ways.

Personability

A friendly and engaging personality is helpful to constructing good personal relations with one's collaborators, as well as to extending one's own network to include useful acquaintances (such as, in the case of Chris Somerville, the friendship of James Watson, whose sympathy proved crucial to starting Arabidopsis research in the US on a massive scale). The ability to establish personal networks is useful in both competitive and collaborative research environments:

Clarity in Exposition
This is a skill that any reader of this thesis will find extremely useful: if I am unable to express my thoughts and motivate my actions in a clear and intelligible way, my chances of convincing you of my arguments are certain to diminish (if not to vanish entirely). The same is true in scientific discussions and writing. Clarity in exposing one's ideas is a means towards the acquisition of authority, especially in the context of gatherings such as the yearly Arabidopsis Meeting, where hundreds of Arabidopsis researchers strive for their peers' attention in order to advertise and discuss their ideas. This skill is also crucial in scientific debates and controversies, where problems are often due as much to ambiguities in premises and methods as they are due to actual dissent. Acquiring this skill is complicated by the predominance of the English language as the official lingo for scientific communication, a factor that certainly favours native English speakers with respect to individuals who have to learn to master a foreign language. The unrivalled status of Anglo-American research as the centre of the Arabidopsis community is no doubt also due to this issue.

Trend-Watching
When selecting research directions, techniques or model organisms, researchers need a sensitivity to upcoming trends both among their peers and among their sponsors. In other words, they need to be able to align themselves with powerful allies, in order to increase their chances of receiving funding or carrying out fruitful collaboration with other scientists. This does not necessarily mean giving up on one's ideas if those ideas do not fit a pre-established paradigm. Rather, it means being able to phrase one's own position and goals in the terms and perspective favoured in the larger community, without necessarily changing the content of one's own proposals (or at least, not radically). This skill is especially evident in the uncanny ability of members of the Arabidopsis community to tap the most sensitive nerves of governmental funding policies: for instance, while the first decades of Arabidopsis research were largely blue-skies, Arabidopsis projects have long been advertised to the NSF as crucial to agronomic development.

A combination of these skills, especially authority and planning, has been necessary to achieving the largely successful enforcement of the ethos of accessibility in the Arabidopsis community. Without the commitment of skilful individual researchers in power positions to enforce collaborative norms, it would have been impossible to establish the ideal to 'share and survive' among researchers used to the competitive strategies promoted in most other big science projects. In this way, the exercise of social skills contributed both to the divulgation of individual understanding and to making available the conditions under which such understanding is achieved in the first place.

7.3 Does Access to Knowledge Guarantee Understanding? A Critique of Critical Contextual Empiricism

7.3.1 Epistemically Acceptable Knowledge

Critical contextual empiricism, or CCE, is one of the labels given by Helen Longino to her views on scientific knowledge. The emphasis on empiricism stems from Longino's long-held naturalist approach, by which she implies treating 'the conditions of knowledge production by human cognitive agents, empirical rather than transcendental subjects, as the starting point for any philosophical theory of knowledge, scientific or otherwise' (Longino 2002, 10). CCE builds on two insights developed by recent work within philosophy, history and sociology of science: the irreducible pluralism of explanatory practices[204] (hence the adjective 'contextual') and the social nature of scientific practices, including theory-making and reasoning as well as experimentation, where 'social' indicates all modes of interactions leading to a revision of the beliefs held by all participants (hence the adjective 'critical'). It is in this context that Longino tries to provide a definition of what scientists should regard as 'epistemically acceptable knowledge':

> Some content A is epistemically acceptable in community C at time t if A is supported by data d evident to C at t in light of reasoning and background assumptions that have survived critical scrutiny from as many perspectives as are available to C at t, and C is characterised by venues for criticism, uptake of criticism, public standards, and tempered equality of intellectual authority (Longino 2002,135).

Longino is here building on the Mertonian tradition of articulating criteria for 'good' scientific practice, while at the same time upholding a Popperian trust in the power of effective criticism to improve scientific results. She proposes four norms that scientists should follow in order to secure discursive interactions and thus produce epistemically acceptable knowledge, all of which stress the conditions under which scientists can communicate (and criticise each other) freely and effectively. The first, *venues for criticism*, points to the necessity of instituting publicly available platforms for scholarly exchanges – such as journals and conferences. The second norm concerns the status and role of critiques in scientific discourse: *uptake of criticism* implies that making and debating critiques should be considered as important as producing new results, so as to force scientists to modify their arguments and results in view of their peers' assessment of them. The third norm speaks for itself: without *public standards*, there would be no way to communicate and, most importantly, assess one's findings in the light of the overall cognitive aims of the scientific community in question. Finally, Longino proposes *tempered equality* of intellectual authority: even in the face of differences among the cognitive capacities, training and institutional resources possessed by each researcher

[204] See Chapter 2, sections 2.1.2 and 2.2.1.

(hence the qualification 'temperate'), there should be opportunities for all members of a given scientific community to express and debate their views and results.[205]

Rather than relying, like Merton, on idealistic motivations, Longino gives a strongly pragmatic flavour to her view: in her eyes, adopting these regulative norms constitutes the only way towards creating reliable scientific knowledge. I here want to point out, however, that her view does not take account of a crucial ingredient for fruitful scientific debate: this is the necessity, for the participants in the discussion, to understand each other at least to some extent. As made clear in the preceding chapters, making the tools and concepts needed for understanding a phenomenon *accessible* to one or more scientific communities is not enough to secure understanding. Scientists need to learn how to use these tools, which in turn implies learning how to interact with members of the relevant communities: namely, there needs to be a constant balancing between accessibility of tools and education about their use and significance in view of the communities' epistemic (and other) goals. Social skills play a crucial role in this context. Possessing the same social skills means being able to communicate with others in the first place; further, the exercise of the right social skills guarantees access to the theoretical and performative tools needed to conduct research on the phenomena of interest.

Further, Longino's disregard for the conditions under which understanding is obtained, and thus for the importance of social, theoretical and performative skills in the social production of scientific knowledge, leads her to overlook an even more urgent problem plaguing communication within and among research communities. This is the idea, illustrated in Chapter 5 and further elaborated in the next chapter, that different researchers might resort to different combinations of tools and skills in order to understand the same phenomena, thus acquiring different understandings of those phenomena. It is difficult to determine whether those different understandings are compatible with each other, and whether reliance on one rather than the other impacts the efficiency of scientific communications. Yet, these difficult questions need to be addressed by scholars who, like Longino, believe that constructive criticisms and dialogue among researchers are what makes scientific knowledge epistemically acceptable. Consider the case of two scientists from two different laboratories, both of which work on Arabidopsis metabolism, meeting up in order to discuss and compare their results. In accordance with Longino's norms, the goal of their exchange is to improve each other's understanding of plant metabolism. Yet, to reach that goal, they first need to acknowledge in which ways their understandings of metabolism differ, as well as the reasons for that difference. I have been arguing that the only way for each of these researchers to understand metabolism in the same way as the other is to acquire a similar set of theoretical and performative skills, which would allow them to perform the same experiments, reason in similar terms and ultimately share the experience of understanding the phenomenon in the same way. The scientific exchange between these two individuals thus presupposes much more than simply the ability to meet and talk to each other: they need to possess the social skills that will allow them to communicate their methods, tools

[205] As Longino notes, the 'exclusion of women and members of certain racial minorities from scientific education and the scientific professions constitutes not only a social injustice but a cognitive failing' (2002, 132).

and skills to their interlocutor, with the goal of enabling her to experience the phenomenon in question in a way similar to their own.[206]

In the next subsection, I shall try to analyse these issues with reference to Arabidopsis research, particularly by focusing on the relation between TAIR curators and the rest of the Arabidopsis community. I intend to demonstrate how Longino's criterion for epistemically acceptable knowledge is insufficient to addressing the problems confronted by TAIR researchers, whose scientific goal is precisely to facilitate epistemic access to knowledge about Arabidopsis. My analysis of communication dynamics between TAIR and its users illustrates that a common, or shared, understanding is very difficult to achieve and assess, thus making of scientific understanding (and the conditions under which it is achieved) a topic of major importance to social epistemology in science.

7.3.2 Reality Check: TAIR and the Arabidopsis Community

Let me start from the observation that the ethos of accessibility characterising Arabidopsis research, and especially the TAIR project, conforms perfectly to Longino's four CCE norms. Concerning *venues,* Arabidopsis researchers are certainly encouraged to frequently attend conferences and meetings in order to hear what others are doing in their fields and present their own work. These conferences are organised on a monthly basis at the national level and of course there is the International Arabidopsis Meeting, which draws thousands of Arabidopsis researchers together every year. Further, the community as a whole places emphasis on building public platforms for data exchange, such as TAIR and other databases, newsletters, Arabidopsis mailing lists, books, journals and, most recently, even blogs. Finally, we have seen how the free exchange of data (also at pre-publication stage), seed stocks and instruments is encouraged in the community. The requirement of *uptake of criticism* also seems to be amply satisfied. TAIR curators are constantly trying to encourage critical feedback from TAIR users (as already noted, one of the curators is actually specialised in education and outreach), as are the stock centres NASC and ABRC. Various groups in the community are also busy trying to establish comparative studies with other plant/animal model communities, thus exposing their work to evaluations by peers from varying social and scientific contexts. As to *tempered equality*, we have seen that the organisation of Arabidopsis research as centralised big

[206] The requirement of at least partial overlap between two researcher's understanding of a phenomenon, in order for them to be able to exchange critiques and information, is in contrast with the description of 'trading' among scientists put forward by Peter Galison. Galison argues that researchers coming from different epistemic cultures do not need to understand each other's view of the phenomena under scrutiny in order to exchange information about them from their respective viewpoints. As an example, he takes the case of experimentalist and theoretical cultures in particle physics: while experimenters provide data to theoreticians, he notes, theoreticians provide experimenters with testable hypotheses; yet, for this exchange to take place, experimenters do not need to understand phenomena in the same way as theoreticians, and vice versa (Galison 1999, 152). I take issue with this description. In the biological sciences, boundaries between epistemic cultures have mostly to do with changes in the way to understand a given phenomenon, rather than (as Galison proposes) with different research goals (e.g. experimental versus instrument-making versus theory-making): and in most cases, researchers need to confront the nature of those changes (thus managing to compare their approach with others) in order to be able to fruitfully communicate with biologists trained in other cultures.

science might pose a threat to the equal distribution of resources and power among participants to the community. However, it is also true that both the MASC and the EU hold regular meetings trying to ensure that all participating groups, also the ones that are not working at the level of molecular or system biology, be granted attention and funding. Finally, there can be no qualms with the way in which the Arabidopsis community creates and enforces *public standards*. The pursuit of integration as a prominent research goal, as well as the elaboration of common terminology and frameworks (such as in GO and PO consortia) and the incentives towards interdisciplinary research make the Arabidopsis community a role model for good planning and successful implementation of standards allowing participants to communicate with each other – certainly in the biological sciences, where the absence of public standards is notorious and conspicuous.

In short, it seems that the Arabidopsis community is doing its best, given its size and breadth of research interests, to comply with the norms of CCE. TAIR, in particular, incorporates and actively tries to respect all four norms. Does this make it into an ideal scientific community, where knowledge is produced via critical confrontation of a plurality of perspectives and is thus, in Longino's definition, epistemically acceptable? I would like to argue that this is not the case, since the contents of their research remain largely uncriticised and unquestioned. As I already pointed out, this has nothing to do with TAIR researchers' eagerness to establish platforms for critical exchanges with their users. The TAIR site offers numerous ways to contact TAIR curators with questions and suggestions. Incoming queries are dealt with swiftly, with replies being posted online so as to be available to anyone interested in the same issue. Yet, the few questions pouring in are ones of clarification or concerning technical details (e.g. 'I want data', 'I don't understand this tool', 'I can't find this tool', 'Please include this info', and so forth), rather than questions concerning the conceptual, interpretive and empirical content of the database.

There are many reasons for the lack of critical attitude among TAIR users. Sue Rhee observes that most researchers are not used to extensive collaboration with others, but rather are trained to work as much as possible with small groups of people, so as to get personal credit for their results.[207] This means that most biologists do not appreciate the ethos driving projects such as bio-ontologies and databases, i.e. the idea to contribute and share data rather than simply profiting from existing resources in order to develop new insight in their own specialised problems. More generally, this can be seen as a fundamental tension between *specialists*, that is researchers focusing their work on very specific phenomena (and thus restricting their skills to the mastery and development of tools and skills relevant only to that context), and *generalists*, that is researchers needing to understand as many phenomena as possible (thus acquiring a diverse set of skills and tools), in order to achieve a more global view of their subject and employ that bird's eye perspective to construct tools towards enhancing integration. TAIR curators, who need to know about several aspects of Arabidopsis biology in order to curate Arabidopsis data, constitute a good example of this second category. However, such a generalist outlook tends to be stigmatised by the broader biological community, where specialist knowledge is held to be of much higher value. This preference is rooted in pragmatic considerations.

[207] Interview held on 17 August 2004.

Publishable discoveries (the ones counting as 'cutting-edge research') are much easier to obtain within very specialised domains than through encyclopaedic knowledge of a series of phenomena. Projects guaranteed to bring swift results within a narrow domain are what counts as cutting-edge research: they receive funding more easily and frequently than projects concerning broad questions that require long-term sponsorship and constitute a much riskier investment, as the proposed goals are more ambitious and thus more difficult to achieve. Consequently, researchers shy away from a generalist training, which is seen as unlikely to attract funding as well as scientific success.

This situation has two main implications. On the one hand, generalist knowledge tends to be pursued by a very small set of researchers, whose work is viewed as a 'service' provided to deserving specialist users.[208] The reliability and accessibility of such services is taken for granted by users, thus making it hard for TAIR researchers to ask for collaboration from the rest of the biological community towards improving and refining their tools.[209] On the other hand, specialists trained to focus on a very narrow set of phenomena lose the ability (in fact, the skills) to critically assess knowledge accumulated by specialists in other fields. It thus becomes difficult for them to assess the research towards integration carried out by TAIR workers, or to reason about a project as ambitious as the GO (which they see not so much as providing the foundations for their own research, but as summarising discoveries that already took place – thus disregarding the impact that such formalisation of biological knowledge could have on their own research[210]). Not only do Arabidopsis researchers lack the willingness to provide constructive critiques to TAIR contents and conceptual framework: they also, arguably, lack the skills enabling them to do so.

Lack of active participation and feedback from TAIR users means that TAIR curators lack sources of criticisms to their work, which, in accordance with Longino's observations, greatly diminishes the epistemic value of the knowledge that they assemble. It also means that TAIR curators take significant decisions on how to portray Arabidopsis research, thus, as I remarked in the previous section, strengthening the divide between

[208] Interview with Doug Becker, programmer at TAIR, held on 18 August 2004.

[209] Interestingly, the Open Bio Ontology projects, including GO, also suffer from such stigmatisation as 'mere services' and are not generally regarded as involving actual biological research (interviews to GO team held in Stanford, CA on 21 August 2004). Sociologist James Evans also regards TAIR scientists as outcasts from the larger biological community, precisely for the generalist, service-related nature of their research (pers.comm.): TAIR members themselves acknowledge that such stigmatisation will make it hard for them to get research jobs after the TAIR project is completed. Remarkably in terms of gender (in)balance, the great majority of researchers currently busy in curating databases and bio-ontologies for model organism communities are women: a situation that might indicate the persistence of gender stereotypes within scientific research, with women more willing than men to take on jobs that are indispensable to the community, yet bringing little recognition or rewards.

[210] Recall from Chapter 5 that TAIR data are largely gathered through compiling and organising data from existing published literature and protocols. This means that curators take great liberty in interpreting and selecting data that they deem appropriate to describe specific phenomena. Arabidopsis scientists do not recognise the extent to which data are interpreted and manipulated within TAIR (through what I called 'vision' in section 5.2.1): they use it as a neutral source of data, rather than actively contributing to the database by verifying the accuracy of the published data, the credibility and usefulness of the visualisations offered by TAIR curators and the possible improvements on the dataset.

central loci of Arabidopsis research and peripheral locations where such decisions are accepted rather than contested. Paradoxically, TAIR risks to become an 'epistemic elite' with monopolised and unchallenged control over the structure, content and functioning of what is tacitly recognised as 'common knowledge of Arabidopsis biology' by the whole community.[211]

TAIR researchers are actually well aware of the risks posed by the lack of interaction with much of the Arabidopsis community. They attempt to correct this tendency through two main strategies: (i) making the representations and practices that foster the understanding - and thus, the further development - of the database and its conceptual basis accessible to all participants; and (ii) establishing measures to check that a minimal level of understanding of how TAIR works and on which assumptions is actually achieved among TAIR users. Among the steps taken to pursue strategy (i), we already found:

- the development of guidelines, freely available online, illustrating how TAIR works;
- collaboration with stock centres to make specimens of Arabidopsis easily retrievable;
- publication of proceedings from GO meetings and curators' meetings, so as to make deliberations on terminology, parameters and family relations to be adopted as accountable and transparent as possible;[212]
- collaboration in the development of The Arabidopsis Book (overview of theoretical developments across local research contexts; see section 4.3);
- small expert meetings (e.g. GO content meetings);
- choosing curators that possess a variety of different specialist understandings, while keeping in mind 'what the user wants'.

The realisation of strategy (ii) is more complex, since it involves using monitoring systems and communication channels that simply do not yet exist between Arabidopsis researchers and TAIR researchers. TAIR curators are currently working on alliances with major journals in plant biology, which would make gene annotations compulsory for submissions (just like keywords) and thus alert researchers to the importance of classifying their work in the terms chosen by TAIR. Also, some members of the team are busy with special training programmes on bioinformatics, annotation procedures and GO, as well as a campaign for inserting bioinformatic skills in educational programmes of BSc curricula in biology.

I would argue that Longino could learn much from TAIR scientists' concerns. TAIR curators are aware that following Longino's four norms is not enough to secure epistemically acceptable knowledge, because they witness how, despite the fulfilment of these norms within the Arabidopsis community at large, constructive communication and critical exchange among Arabidopsis researchers is still scarce – especially when it comes to the tools, models and concepts that they all have to agree upon in order to carry

[211] This is especially true since, as we saw, TAIR members collaborate closely with (and are based in the same location as) some of the most powerful members of the MASC.
[212] This archival documentation is available for free at www.biocurator.org.

our their research.[213] Longino's framework thus fails to confront a crucial issue in social epistemology: that is, 'the problem of common representation in diverse intersecting social worlds' (Star and Griesemer, 1989). Longino's failure represents an important finding for my current purposes. In my view, the knowledge produced by the Arabidopsis community can be deemed as epistemically acceptable (as in Longino's definition) only if its members are both able and willing to critically assess the body of knowledge that the community produces, in light of its epistemic goals, as well as the role of their own contribution towards such achievement. TAIR researchers have long noticed that, in order for researchers to effectively critique each other's work and thus improve the quality of knowledge in their community, they need to be able to understand each other. Such understanding can only be acquired through recourse to similar skills and experiences: this means that participants in a research community, no matter their specialisation, should be given a minimal set of skills (in the form of training and exposure to diverse research activities) for understanding their own research in the light of the research efforts and results of the whole community. Researchers need to be trained in ways appropriate to facilitate their dialogue with members of epistemic cultures different from their own: the possibility of establishing such common understanding depends greatly on the social skills of individuals, since it is through social skills that they acquire the performative and theoretical skills that will enable them to experience phenomena in ways similar to their peers. The 'share and survive' motto of Sue Rhee should not imply a mere commitment to making all data freely and easily accessible; it should also imply a commitment to train as many scientists as possible to make the best of those tools and data. This is an insight that has been picked up and emphasised in TAIR itself; yet, it receives almost no attention in Longino's framework, which is arguably too far removed from scientists' practices to take account of the work and skills needed to acquire understanding, as well as the differences in understanding that could derive from differences in skills, experiences and labour distribution within any one scientific community.

Ultimately, Longino falls into the same trap in which Hempel fell half a century ago: she does not acknowledge that the mere availability of knowledge does not guarantee understanding. Her choice of norms enforces openness and critical exchange among scientists: she does not, however, have anything to say about the manner in which such exchanges should take place. She does not show interest in whether researchers are talking past each other or employing standards, skills and tools so different that communication, even if it exists, does not really lead to acquiring a better awareness of how others understand phenomena: what matters in her framework is that researchers can physically meet, spend their time in conversations or in studying other people's work, and confront their peers if they feel that they should. Longino does not say anything about the quality of communication among scientists: in particular, she does not talk of

[213] Longino might object that the communication problems encountered within the Arabidopsis community can be accounted for within her view, insofar as they are due to disregard for the norm 'uptake of criticism'. However, Longino's definition of uptake of criticism does not set any criteria for what counts as a valuable criticism and, even more important, for what counts as a valuable form of uptake. Thus, in a superficial sense, Arabidopsis researchers show respect for the norm by constantly exchanging information and modifying their work accordingly. Yet, this type of uptake is not necessarily based on a shared understanding and thus does not necessarily lead to the wished-for meaningful critical discourse.

the impact that diverse research conditions might have on each individual's understanding of the same phenomena. I see this as a major omission. In order for these norms to work according to Longino's wishes, researchers need to respect a basic requirement: they need to try to acquaint themselves with the skills and knowledge used by others in order to understand a phenomenon. To criticise each other's views, they have to acquire an awareness of how the other understands phenomena: only on that basis can researchers compare their understanding and the other's, assess whether they are in any way compatible and to which extent the differences in results are dictated by differences in the conditions under which they carry out their research.

Chapter 8. Committing to understand

The recognition of a person in the performance of a skill or in the conduct of a game of chess is intrinsic to the understanding of these matters. We must surmise that we are faced with some coordinated performance, before we can even try to understand it, and must go on trying to pick out the features that are essential to the performance, with a view to the action felt to be at work with it

Michael Polanyi 1966, 30

8.1 **Scientific Understanding in Biology (and Beyond?)**

This final chapter has the predictable function of bringing together the various threads of my analysis of scientific understanding. Less predictably, I also aim to highlight some implications of my views, including the relation between scientific understanding and other types of understanding, as well as between understanding and knowledge (section 8.1.1); the extent to which my perspective is compatible with the one offered by a prominent contributor to the contemporary debate on scientific understanding, Henk de Regt (section 8.1.2); and the ways in which some of my results could be developed as a contribution to biology itself, in the 'complementary' fashion proposed in Chapter 4 (section 8.4). The most important goal of this chapter, however, is to elaborate on the philosophical significance of my analysis of biological understanding as achieved by Arabidopsis researchers. I develop a novel perspective on the relation between epistemic skills and what I call research commitments (section 8.2), which in my view helps to clarify the differences among three main kinds of scientific understanding, which I identify in section 8.3 as theoretical, embodied and integrated understanding. Particular attention is devoted to integrated understanding and its role in contemporary model organism research (section 8.3.2).

8.1.1 The Result of a Well-Coordinated Performance

Lack of clarity about the nature of understanding and the conditions under which it is acquired generates confusion with regard to the differences separating the notion of understanding from the one of explanation. As I outlined in Chapter 2, many philosophers of science believe these two notions to be epistemologically equivalent. Consider for instance the tentative definition of scientific understanding provided by Lacey in 1999: ‘understanding reality involves grasping the ‘what?’, the ‘why’, and the ‘what possible?’ of phenomena’ (95). Lacey develops his claim through an analysis of the nature and possible answers to ‘what’, ‘why’ and ‘what possible’ questions: he then investigates the types of explanations through which understanding is achieved and assumes, as many other philosophers have done, that an investigation of explanatory strategies will lead to a philosophical understanding of how scientists understand natural phenomena.

I see Lacey's strategy as misguided, as it focuses on the *objects* and *results* of understanding rather than on the *process* of understanding as such and the *conditions* under which it takes place. This does not mean that the nature of the phenomena to be understood, as well as of the tools, theories and models used to gain such understanding, do not impact the quality of the understanding that is obtained: in Chapters 5 and 6 I have examined precisely how models and theories are used as tools to understand phenomena – and how, depending on the selected combination of tools, scientists' understanding of the same phenomenon might change. Rather, I am claiming that Lacey's approach does not consider a crucial philosophical issue in the study of understanding, that is, the meaning of the term 'grasping' used in his definition. Regardless of the specific nature of the answers given to scientific questions, what is it that makes us able to grasp their relevance to our understanding of phenomena?[214] My analysis of understanding as obtained by (plant) biologists tried to address precisely this question. I am now in a position to discuss my findings and their epistemological implications in a systematic fashion, thus sketching a philosophical account of the nature of scientific understanding by reference to the conditions under which it is acquired in the life sciences.

In my view, scientific understanding is a cognitive achievement resulting from performing a series of actions (including both mental actions, such as thinking, and bodily actions, such as ways of moving and interventions on phenomena). The performance of such actions needs to be skilful: biologists seeking understanding of a natural phenomenon need to master theoretical, performative and social skills that will help them to adequately reason about, interact with and share knowledge about that phenomenon (often through the correct handling of theories, models and instruments). This is because the exercise of skills enables biologists to acquire and coordinate theoretical and embodied knowledge about the phenomenon under scrutiny in a manner that is adequate to the material, scientific and social environment in which they find themselves. Such coordination is what ultimately grants a scientific understanding of the phenomenon. Knowing a theory about a phenomenon does not suffice to understand it, as understanding springs from knowing how to apply such theory to the phenomenon and thus use it as a platform for further investigations. Similarly, knowing how to manipulate some material features of a phenomenon (such as a Columbia specimen in the case of plant biology) does not imply having an understanding of it. At least some theoretical knowledge, including concepts appropriate to the categorisation of and reflection upon sense-perception and data gathered through measuring instruments, is needed to systematise and order information accumulated through interaction with the phenomenon as well as to recognise how previously accumulated biological knowledge applies to that phenomenon.

The prominence of scientists' skills and experiences in achieving such coordination between thoughts and actions, conceptual analysis and sense-perceptions, effectively separates the understanding thus obtained from the mere possession of theoretical knowledge in the form of an explanation. Theoretical knowledge, per se, does not provide biologists with the theoretical and performative skills needed to trace its significance for

---

[214] To Lacey, 'grasping' means 'possessing an account of': his account thus falls in the category of philosophical positions trying to reduce understanding to explanation (see section 2.1.2).

natural phenomena. This is the sense in which, as I stated in my introduction, the possession of theoretical knowledge in the form of theories or explanations does not automatically imply the ability to use such knowledge to understand phenomena. In fact, I would like to push my argument further: in science, to 'know' implies not only to possess theoretical knowledge, but also to be able to use it towards understanding actual phenomena. The ability to use theoretical knowledge to that aim is granted by embodied knowledge, and particularly by components of embodied knowledge that I called epistemic skills. Possessing skills relevant to coordinating theoretical and embodied knowledge of phenomena is thus a prerequisite for an individual to declare that he or she 'knows' something about such phenomena: scientific knowledge without understanding amounts to a mindless reading of unintelligible concepts and data, and should therefore not be considered as adequate knowledge.[215] In short, understanding constitutes a precondition for the acquisition of scientific knowledge: it is not acceptable for scientists to declare that they know something about a phenomenon (for instance, an explanation of a behaviour or a physiological process displayed by a specific organism), unless they possess skills enabling them to understand that phenomenon in the light of such knowledge.

The skilful coordination of theoretical and embodied knowledge is crucial to understanding phenomena in two ways: (1) as I illustrated when analysing modeling practices in Chapter 6 (section 6.4), it allows biologists to select beliefs, thought processes and experiences that are *relevant* to the phenomenon in question and (2) it enables biologists to successfully *integrate* these components with the goal of applying them to the phenomenon.[216]

Let me exemplify how my views apply to the case of a biologist seeking to understand a specific phenomenon in Arabidopsis biology: for instance, the phenomenon of growth regulation, that is the process by which a plant 'decides' when and in which size to develop roots, stem, leaves and (if at all applicable) flowers and fruits. Attempts to understand the complex processes underlying growth regulation in plants have been at the centre of (among others, Meyerowitz's) Arabidopsis research since its inception. This is hardly surprising: making sense of how plants germinate seeds, metabolise nutrition and develop leaves, flowers and roots is essential to being able to intervene in their

[215] Note that this approach is different from the definition of knowledge as 'justified, true belief' that stands at the centre of many an epistemological discussion. An in-depth analysis of the implications of my argument for this view would require a second dissertation and I will therefore not pursue it here. I shall only point out that, while not necessarily incompatible with it, my definition of adequate knowledge differs from the standard definition of knowledge as justified, true belief in two main respects. On the one hand, it highlights the difficulties in establishing what amounts to a 'justified' belief in science (a feature that is not emphasised within general epistemology, where the actual conditions under which scientific beliefs can be considered as justified are rarely the object of critical scrutiny). On the other hand, it emphasises the experience and abilities of the knower, as well as the circumstances and aims towards which knowledge is acquired, as important characteristics of the *nature* of scientific knowledge, rather than just of its context of application. In this sense, my approach follows in the steps of the early pragmatist tradition.

[216] Galison uses the expression 'coordination of action and belief' to describe the interaction between experimentalist and theoretical culture in physics (1997, Chapter 9). The practices to which he refers constitute, in my eyes, an excellent case for the analysis of understanding that I propose here, even if my emphasis on theoretical and embodied knowledge is largely unrelated to Galison's account.

development and modify it as required by private or public sponsors. Current research has acquired a good level of knowledge especially concerning the molecular bases for this phenomenon: that is, the biochemical pathways through which signals regulating growth are sent to responsible organelles throughout any given plant. Accessing such knowledge is relatively easy. Our biologist needs only to open a textbook on plant development to find that much of growth regulation takes place thanks to one single gaseous hormone called ethylene. Ethylene acts as an endogenous regulator of plant growth and development via a process called 'ethylene signal transduction'. Thanks to this process, germinating seeds (grown in the dark) become sensitive to the level of ethylene in the environment, which can change dramatically depending on environmental conditions. Plants learn to respond to ethylene levels with a range of inhibiting or stimulating responses. Thus, the ability to measure ethylene levels becomes a way for plants to identify and prepare for all sorts of environmental stresses, from pathogen attacks to flooding. Further, specific quantities of ethylene also signal the presence and eventual competitive advantage of other plants in the neighbourhood (for instance, whether there are very tall plants nearby, which are therefore likely to steal all the sunlight): the plant is therefore able to regulate the timing as well as the size of growth processes in accordance with environmental variations (e.g., plants would not bear flowers in very low temperatures and would develop strong stems and roots if they perceive their environment to be particularly competitive).

Now, it may seem that this explanation of growth regulation enriches one's understanding of this phenomenon without recourse to specific background knowledge and skills. Even readers who have no training in biology should have understood something about growth regulation from reading the above paragraph. However, what kind of understanding is this? I would suggest that the understanding gathered thanks to the above explanation is extremely superficial and vague, unless: (1) we are aware of the definitions attributed by plant molecular biologists to the terms 'response', 'transduction signal', 'competitive advantage' and 'biochemical pathways'; (2) we know what parts of plant morphology are identified via the terms 'stem', 'root' and 'seeds', as well as which parts of the plant are involved in realising the process of signal transduction; (3) most importantly, we are able to bring together our awareness of how the plant is structured with the concepts and processes specified above, thus effectively understanding how precisely ethylene signal transduction takes place and what are its effects on the plant. Rephrased in the terminology that I introduced so far, point (1) represents what I call theoretical knowledge; point (2) corresponds to embodied knowledge about the phenomenon at hand; and point (3) refers to the 'coordination of theoretical and embodied knowledge' that I indicated as an essential condition for understanding the phenomenon.

What about skills? In my example, the use of skills already comes to the fore: it is clear that, in order to be able to use the relevant knowledge, we must possess theoretical skills enabling us to reason through theories about mutation responses and transduction, as well as performative skills deriving from interaction with the plant and, possibly, the handling of illustrations (pictorial models) representing the process of ethylene reception. Yet, the importance of exercising *specialised* skills, which can only be acquired through scientific training and research experience, is not obviously implied here. This is actually because the explanation that I gave above is not the explanation of growth regulation that you would find in a basic biology textbook: it is my ‘translation’ from biological jargon to something closer to everyday terminology, which I carried out precisely to allow non-biologically-trained readers to acquire some common-sense understanding of the phenomenon that the explanation is meant to illustrate. Contrast my simplified explanation of the role of ethylene in growth regulation with the following excerpt from an introductory chapter in The Arabidopsis Textbook, summarising recent findings on ethylene in what biologists would regard as extremely simple terms:

> Ethylene affects many processes throughout the plant's growth and development, but most of these effects are cumbersome to use for a mutant screen. For example, although ethylene modulates the timing of leaf senescence in Arabidopsis, it is not required for the senescence syndrome to occur. Thus, mutant isolation in Arabidopsis has relied almost exclusively upon one mutant screen: the triple response. Dark-grown seedlings exhibit several phenotypic responses to ethylene that are collectively termed the “triple response’. The triple response in Arabidopsis seedlings is characterized by a shortened and thickened hypocotyl, an inhibition of root elongation, and the formation of an exaggerated apical hook. These features contrast sharply with the etiolated phenotype observed in dark-grown seedlings exposed to air. The readily distinguishable phenotype and the ability to screen thousands of seedlings on a Petri dish have greatly facilitated the identification of mutants that affect ethylene signalling in Arabidopsis. Mutations isolated using a screen for an altered triple-response to ethylene fall into two main classes: (1) mutations that render a plant insensitive to ethylene; and (2) mutations that result in a constitutive ethylene response. Ethylene-insensitive mutants display a similar phenotype when grown in ethylene to what they display in air. Examples of ethylene-insensitive mutants are *etr1-1*, a gain-of-function mutation in an ethylene receptor, and *ein2*, a loss-of-function mutation in one element of the signal transduction pathway. Constitutive ethylene-response mutants display a triple response in both air and ethylene. A constitutive ethylene response can result from ethylene over-production (*eto* mutants). Alternatively, mutations in the signal transduction pathway can also lead to a constitutive ethylene response as has been found with the *ctr1* mutant. (Shaller and Kieber 2002, 1)[217]

[217] For reasons of space, I have eliminated all references to landmark papers that peppered every line of this short excerpt (e.g. ‘[Kende 1993]’). I eliminated 10 references in total. This does not mean that the availability of references is not relevant towards the biological understanding of growth regulation to be acquired through reading this paragraph. A biologists reading it is likely to possess theoretical, performative and social skills allowing him to relate the text to the actual discoveries and experiments made by each of the persons cited, thus greatly enhancing the theoretical and embodied knowledge that the biologist coordinates towards understanding the phenomenon. The ability to interpret references is thus per se a sign of the gulf between the understanding of plant growth regulation acquired from such a paragraph by a trained biologist and the (little) understanding acquired by a layperson with no adequate epistemic skills.

In my view, understanding the phenomenon of plant growth regulation through reference to this passage implies a series of theoretical and embodied skills that can only be acquired through training in molecular biology as well as experience in experimental research on Arabidopsis signal transduction pathways. I defy a reader who possesses none of such skills to understand this phenomenon to a level acceptable in order to be able to communicate with biologists, further research on this topic and manipulate plants according to this knowledge. How would this be possible, unless one knows what a mutant screen is, what it looks like and how it is isolated experimentally; how to observe and evaluate the 'formation of an exaggerated apical hook'; or how to distinguish an etiolated from a non-etiolated seed phenotype among the hundreds visible from a Petri dish? In fact, I readily admit to not possessing such an understanding myself, despite the time spent in Arabidopsis laboratories and studying Arabidopsis papers: my theoretical and embodied skills, as well as my background knowledge about plant biology, are not enough for me to be able to use information such as the explanation above to understand the phenomenon of plant growth regulation in the same ways as trained biologists would. I would not be able to use my understanding of this phenomenon to investigate it further and share my findings with other researchers: in short, my understanding is not scientific, as it does not allow me to contribute to biological research on the phenomenon of interest by employing biologists' skills, methods and knowledge.[218]

This example shows that, as I specified in the last chapter, there is a difference between understanding a phenomenon scientifically and understanding it in other ways: a scientific understanding of a phenomenon presupposes acquaintance and mastery of specialised tools, skills, concepts and methods that are used within scientific communities producing knowledge about that phenomenon. Further, this example illustrates how the ability to understand a phenomenon is crucial to the claim of possessing knowledge about that phenomenon. This is a useful reminder that much of science, and certainly much of biology, is concerned primarily with understanding phenomena, rather than with producing theoretical knowledge in the form of theories and explanations. Theoretical knowledge is an indispensable tool towards acquiring understanding of phenomena: it is not an end in itself, nor is it a sufficient tool to that aim - since, as we have seen, it needs to be skilfully coordinated with embodied knowledge in order to provide such understanding.

Note that this conclusion opens up several interesting questions concerning the status of scientific understanding with respect to other types of understanding obtained through the exercise of different skills (such as philosophical, political or literary understandings of phenomena). For instance, does the fact that I do not understand growth regulation in a scientific manner imply that my understanding of this phenomenon is wrong, superficial or useless – or that I, as a philosopher, would not be able to contribute to a scientific

[218] It could be argued that my familiarity with biologists' tools and terminology allows me to gain a partial or incomplete scientific understanding of the role of ethylene in plant biology. Such partial understanding may in fact be viewed as a necessary condition for performing complementary science. Yet, here I do not wish to focus my arguments on the degrees of scientific understanding (e.g. poor or good) possessed by any one individual; rather, I want to focus on the qualitative differences between kinds of understanding that can be found both within science and between scientific and other human activities.

discussion on these findings by employing skills differing from the ones adopted by biologists? I do not think so, as clearly indicated by my stance on complementary science (Chapter 4). What I want to point out is that the scientific understanding of a phenomenon is the result of the correct application of a number of skills, depending on the scientific context to which such understanding is supposed to contribute. This does not make other types of understanding, obtained with other types of skills, useless or irrelevant to science. Indeed, science might be argued to represent only one of many ways in which the same phenomenon can be understood: which type of understanding (e.g. scientific, political, aesthetic, popular, literary) is most helpful depends on the context in which the phenomenon is being analysed, as well as the purposes of the analysis. For instance, there is no doubt that my scientific understanding of growth regulation is very scarce compared to the scientific understanding obtained by a trained biologist specialised in this field of inquiry; as a result, it might be claimed that I do not have much knowledge of this phenomenon. On the other hand, my philosophical skills enable me to reconstruct the methods by which such scientific understanding is achieved. My philosophical understanding of Arabidopsis research hopefully sheds light on aspects of such research that biologists tend not to think about, or anyhow not systematically. I might thus contribute something to science even without possessing a scientific understanding of phenomena. Further, each individual might possess skills that allow him or her to understand the same phenomena in different ways. Hence, a biologist is not condemned to a 'purely' scientific understanding of a certain phenomenon simply because she possesses and exercises the relevant skills: she might also be able to understand her own practice philosophically and refine this second type of understanding by discussing and reading philosophy.

A second sphere of inquiry triggered by my position concerns the nature of scientific knowledge as such. If it is indeed the case that understanding is a prerequisite for possessing knowledge, can people without scientific training ever gain any knowledge of nature through science? This question brings me back to an issue I raised already in Chapter 2, concerning the internalist nature of my account. To which extent do we need particular kinds of expertise in order to understand and evaluate the results of scientific inquiries? In order to reflect on this question, I propose that we distinguish the types of understanding (and knowledge) needed to contribute to science itself from the types of understanding (and knowledge) needed to assess the significance of scientific (and technological) results for sociological, political, economic or theological issues. In the latter case, more than a scientific understanding of the phenomena is needed: for instance, an individual might understand how a plant is genetically modified, yet not how such intervention impacts on national economy or local ecology once GMOs are cultivated on a large scale. To assess the social implications of scientific findings, a variety of understandings is required of which scientific understanding is only one type: there is thus a wide space for scientists and non-scientists to interact and exchange expertise and opinions. The case of understanding aimed at contributing to science is different. It could be argued that such understanding is, by and large, scientific, as it is achieved mostly by individuals trained and socialised to study the phenomena under scrutiny. This does not imply that individuals who do not possess scientific understanding of specific phenomena are entirely unable to contribute to scientific research on those phenomena. Take, for instance, current investigations of avian influenza, encompassing its actual and potential distribution and the probability of human contagion. Understanding the features of the virus, and thus deliberating on the implied health risks for humans and other animals, requires skills such as bird-watching and knowledge concerning the behaviour and movements of birds in the wild and in (not entirely industrialised) farms. These skills and knowledge are not acquired through scientific training, but through years of hands-on experience in farming and/or bird watching – and in fact, scientists investigating avian influenza capitalise on the contribution of farmers and bird-watchers to their investigations. These contributions by non-scientists are certainly valuable to scientific research; yet, they can arguably be dubbed as scientific knowledge only once they are tested and extended through scientific methods, as well as integrated with theoretical and embodied knowledge of viruses and disease.

A full discussion of this topic and its implications would require more than a mere dissertation. What I want to do here is simply to point at the range and potential significance of philosophical questions emerging from my views on understanding.[219]

8.1.2 Similarities and Contrasts With De Regt's Views

As already indicated in Chapter 2, my approach to scientific understanding in biology owes much inspiration to Henk de Regt's research on understanding in science (1999, 2001, 2004, 2005). For a start, we are both interested in understanding as expressed and sought for by practicing scientists within specific research contexts and with the help of available tools (as de Regt notes, there is 'a variety of toolkits for understanding'; 2001, 261). In fact, we both oppose what de Regt calls 'objectivist' accounts of understanding, in which 'one understands a phenomenon if one possesses all relevant knowledge' (de Regt 2004, 100). Additionally, I share de Regt's emphasis on the pragmatic side of understanding and his stance that such pragmatism should not be confused with subjectivism (or, in the words of Trout, with a 'feels right' experience of 'intellectual satisfaction'; Trout 2002, 213): rather, it indicates the extent to which scientific understanding results from researchers' ability to use and apply the theoretical knowledge that they possess.

There are also ways in which my approach and de Regt's differ considerably. In many respects, this is a matter of emphasis. For a start, de Regt is interested in scientific understanding as an aim of science, rather than as a cognitive process. These two readings of the same phenomenon are by no means incompatible: as made clear in the previous section, understanding nature is both an aim and an achievement to practicing scientists. Also, de Regt does not discuss the difference between individual and social understanding, nor does he analyse the extent to which scientific understanding depends on social sharing and communicative skills: still, paying attention to such topics represents a continuation of, rather than a departure from, his approach. The same holds for de Regt's emphasis on intelligibility as an epistemic value with which to assess

[219] More on the prospective implications of my views can be found in section 8.4.

scientific theories (2001, 261): in my present work, I have simply avoided to discuss intelligibility, as such discussion was not relevant to my purposes.

I now wish to focus in more detail on the aspects of de Regt's view of which I am outright critical. All of them concern the primary role taken by theoretical knowledge in his account, at the expense of the type of knowledge that I call 'embodied'. De Regt insists on theories as a necessary means towards acquiring scientific understanding, as illustrated in his Criterion for Understanding Phenomena (2005, 150):

> *CUP: A phenomenon P is understood if and only if a theory T of P exists that is intelligible (and meets the usual logical, methodological and empirical requirements).*

As made clear in the previous chapters, understanding in biology does not necessarily have to be obtained through theories: biologists understand phenomena also by building and handling models of those phenomena as well as by exploring and interacting with the phenomena themselves (e.g. Chapter 6). Further, even when understanding is obtained through reference to a theory, it is important to specify that there are many types of theories available in the life sciences, each of which has to meet different 'logical, methodological and empirical requirements' depending on the context in which it is used (see Chapters 2 and 5). I think that de Regt should expand his account of the role of theories in acquiring understanding in the light of these objections – or, more simply, delete the 'only if' clause from his CUP, thus recognising the role of other components of scientific practices in providing understanding.

A second problematic aspect of de Regt's conceptualisation consists in his definition of intelligibility, as expressed in his Criterion for the Intelligibility of Theories (2005, 151):

> *CIT: A scientific theory is intelligible to scientists S (in context C) iff they can recognise qualitatively characteristic consequences of T without performing exact calculations.*

While agreeing with de Regt that 'intelligibility depends not only on the virtues of T itself but also on such contextual factors as the capabilities, background knowledge and background beliefs of the scientists in C' (2005, 151), I think that his distinction between qualitative and quantitative aspects of theories, as well as of their consequences, does not at all respect this crucial premise. De Regt seems to equate 'qualitative consequences' with aspects of a theory that are intuitive to scientists using it: as he specifies in a footnote, this means that such aspects do not require 'complete logical argumentation' in order to become clear to scientists (2005, 167). This characterisation of intuition does not take account of the *differences in expertise and skills* among scientists working in different fields. Scientists who possess strong theoretical skills, such as the ability to work with specific concepts and symbolic notations, might find the quantitative implications of a theory (and/or logical argumentation supporting them) to be more intuitive than its qualitative consequences. Scientists working in applied areas, such as engineering and experimental biology, might find calculations to provide more understanding than their own intuitions about the qualitative consequences of a theory, *depending on the phenomenon* that they are analysing. For instance, it could be argued that areas such as evolutionary-developmental biology, nanotechnology or quantum computing require scientists to disregard their qualitative intuitions and develop quantitative treatments of phenomena, since qualitative intuitions about implications of the theory for the phenomena at hand are misleading and confusing. In fact (and this is where I think de Regt's account needs substantial amendments), nowhere in de Regt's definition is it explicitly specified that such 'qualitative consequences' of a theory consist of the relation and/or significance of the theory with respect to the phenomena to which it can be applied. From de Regt's examples from the history of science, it becomes obvious that this is what he means. I think, however, that such a relation should explicitly feature in his definitions of understanding and intelligibility, as establishing a relation between theories and phenomena is crucial to the ability of researchers to use theories to understand phenomena. As I illustrated throughout my thesis, understanding a phenomenon results from coordinating theoretical and embodied knowledge that is *relevant* to that specific phenomenon: there is little in de Regt's account that helps us to identify relevant knowledge and/or to analyse the relation between the intelligibility of a theory and its potential applicability to the phenomena to be understood. In the CUP, empirical requirements do not figure as major constraints on the achievement of an intelligible theory, but are rather mentioned as a mere addition to intelligibility. While carefully analysing the role of theories in the acquisition of understanding, de Regt seems to take phenomena for granted.

Against my critique, it could be argued that the differences between my approach and de Regt's spring from differences in the objects of our analyses, that is, the physical sciences (in his case) and the biological sciences (in mine). Whether this is true is of course a very interesting question, especially given the many existing claims about biology being more disunified and pluralistic (in its theories, models and perspectives) than the physical sciences.[220] I have myself proposed, in Chapter 2, that biology has a dual nature, i.e. that theoretical knowledge in biology can never be dissociated from or favoured over embodied knowledge. This feature might be taken as a reason to emphasise embodied knowledge and skills in the context of biological research, without necessarily being able to apply the same approach to the physical sciences. This would mean that de Regt is right in emphasising theoretical knowledge as a principal means towards understanding in the physical sciences, while his account does not adapt as well to sciences where theories, and particularly formal theories expressed in mathematical terms or strictly logical connections, play a less central role. Is it then the case that biological understanding differs substantially from understanding in the physical sciences?

I am actually rather sceptical about the supposedly 'special' features of the biological sciences and of the understanding therein acquired. Notably, when arguing about the dual nature of biology I have not insisted that such a framework would not hold for other

[220] This view of the difference between biological and physical sciences is implicitly supported by many of the philosophers working on disunity and pluralism in biology. Explicit claims in this regard have been advanced, among others, by Mayr (1982, 2004); van der Steen (1991); Wimsatt (1972); Cummings (1983, ch.1) and Haugeland (1998[1978]).

sciences. In fact, I think that a close look to the practices characterising other fields is likely to prove that the physical sciences are (1) as disunified and pluralistic and (2) as dependent on embodied knowledge as biology is.[221] I have no firm grounds on which to defend this intuition: much further research is necessary in order to explore the applicability of my views to other fields and the way in which the organisation and knowledge produced in other sciences resembles the situation that I ascribed to the life sciences in this thesis. In the absence of such research, the eventual applicability of my views on understanding to sciences other than the biological ones (and thus, the extent to which my views and de Regt's on the role of theories actually clash) remains an open question.

8.2 Three Types of Skills, Three Types of Commitments

8.2.1 Skills and Commitments

As illustrated in the last section of Chapter 6, the exercise of epistemic skills allows researchers to identify which parts of the (theoretical and embodied) knowledge that they possess are relevant to the investigation of the phenomenon under scrutiny. Here I want to discuss the further claim that the skilful performance of research practices both requires and enables scientists to recognise specific bits of knowledge as a stable background to their investigations. Such recognition can be self-conscious. For instance, a researcher might explicitly recognise that his skills in manipulating whole plants are higher than his skills in handling microscopic parts of the plants, such as DNA samples (or vice versa); he will therefore gear his experimental activities to a field of inquiry allowing him to use his skills, such as ecological or physiological research, rather than to an area where he is sure to fare badly (such as microbiology). Researchers might also be unaware of how their skills steer their research directions: for example, a cell biologist might learn to search TAIR for Arabidopsis data without acknowledging that this implies subscribing to the gene-centric theoretical framework adopted by TAIR curators to model and display the data.

This apparently trivial point has important epistemological implications. Skills enable researchers to perform a set of activities leading to specific results. As we saw in the examples of TAIR and NASC, skills do not come easy: considerable time and effort is invested in learning a set of skills, certainly in the case of most theoretical or performative skills used in contemporary biology. This means that the set of skills available to any one researcher is limited. Each researcher will bother to learn and perfect skills that, he or she believes, will be helpful in pursuing his or her research interests. It also means that trained researchers, who already master several skills, have a strong tendency to pursue projects where those skills can be exercised, rather than embarking on projects where (a) they will have to invest time in mastering new skills and (b) they will

[221] I take work by Peter Galison, Jordi Cat, Nancy Cartwright and Hans Radder to add empirical significance to my claim here, even if none of these scholars are engaged in a direct comparison between the physical and the biological sciences.

not be able to use their old skills.[222] In other words, through acquiring specific skills, researchers become committed to the actions, concepts and research directions enabled by the exercise of these skills. Whether researchers explicitly acknowledge it or not, certain principles, concepts, ways of acting and handling objects become *entrenched*[223] in the practice of researchers engaged in the study of a specific phenomenon. They represent knowledge that is assumed, rather than hypothesised, to be relevant to the study of the phenomenon in question. When this happens, those bits of theoretical and embodied knowledge start to be regarded as a necessary platform to carry out research.[224] They become what I shall call *research commitments*, encompassing items as diverse as the theoretical perspective held by the individual biologist engaging in research; the research goals and interests in his or her work; the research conducted by his or her research group; gestures and ways of moving; and the assumptions about the representativeness of the research materials on which the researcher works, as well as the applicability of his or her results.

Research commitments are strongly tied to epistemic skills. They are two aspects of the same process: acting skilfully implies committing to the activities and results that those skills bring forth, while making a commitment to a specific technique or concept requires learning skills adequate to follow up on such commitment. Yet, it is clear that the two notions should be clearly distinguished. As we saw in Chapter 5, epistemic skills denote the abilities to perform an action in a specific manner; commitments consist of a tendency towards (or preference for) pursuing goals that can indeed be obtained through performing that action. As I already remarked, skills are costly: the acquisition of certain skills rather than others represents an investment in the goals that those specific skills will enable one to pursue. This means that the possession of a skill entails a commitment to exploiting that skill (thus taking advantage of an otherwise useless investment[225]). Also, once a commitment is made, it implies learning and exercising skills that are relevant to fulfilling that commitment. This is what makes commitments different from a mere promise or a pledge: they imply, and result from, skilful action.

To underline further the link between skills and commitments, I shall now trace a categorisation of commitments that runs parallel to the categorisation of skills proposed in earlier chapters. I propose to distinguish three types of commitments, depending on the type of goals to which they apply. *Theoretical commitments* are commitments to using or investigating specific concepts, theories and principles. Much has been written about these types of commitments in terms of biases, theoretical perspectives and background

[222] This is especially true within current research practices, which are geared towards producing as many short-term results (and publications) as possible. In this landscape, there are no resources (time as well as funding) to allow researchers to acquire new skills without knowing whether those skills will be useful.
[223] See Wimsatt 1986.
[224] Elihu Gerson colourfully describes this situation as follows: 'our tentative alliances [with phenomena] become stronger as they are used successfully as tools and materials in other projects. It is at this point that our theories become 'facts', firm commitments to act in a certain way' (1998, ch.2, 13). Note the similarity to Hacking's formulation of entity realism, where the ability to manipulate an unobservable entity in the laboratory constitutes a criterion for believing in the existence of that entity (1983, 265).
[225] Rouse (1987, 87ff.) provides an interesting analysis of the role of research opportunities, as perceived by researchers, in shaping the directions and content of science.

knowledge. An example of theoretical commitment is the gene-centred perspective adopted by TAIR as a foundation for its modeling practice. *Performative commitments* consist of commitments to specific habits, that is, ways of thinking or moving that become entrenched to a scientist's practice. For instance, some actions performed by TAIR curators (such as switching on and off a computer or doing a search on TAIR databases) constitute performative commitments by virtue of their being (after years of practice) habitual, unchallenged practices – yet indispensable to carrying out research. *Social commitments* apply to the interests and values endorsed by scientists as a consequence of their financial, professional and personal dependence on specific social hierarchies, sponsors or practices. They affect scientists' epistemic resources, without however being a matter of epistemic dependence themselves. For example, peer review mechanisms engender a series of commitments towards wedding one's work to recognised keywords, the expectations of prospective peers and the content of journals in which one hopes to publish; funding agencies might require scientists to observe specific criteria and/or values in their research, such as, in the case of corporate sponsors, the search for commercially applicable ways to mutate plants and a predilection for short-term results over long-term research projects; or, scientists working in a university or laboratory dominated by powerful deans and supervisors might have to adopt these people's views (including their theoretical and embodied commitments, but also their preferences for values and directions in research) in order to be granted employment or funding for their research.

Commitments have varying time-scales and contexts: a researcher can commit to using a specific technique in order to participate in a given research project, but might never use that technique again once that project ends; on the other hand, a biologist can commit to using animals in experiments and hold on to that commitment for the rest of his career. Peter Galison exemplifies this point by distinguishing between *long-term commitments*, which he describes as 'metaphysical' and which concern methods, principles and goals transcending belief in specific explanations or theories; *middle-term commitments*, that is 'programmatic goals or laboratory practices'; and *short-term commitments*, shaped by particular theories and models that are constantly subject to changes and revisions (Galison 1987, 244[226]). What I find especially interesting in Galison's distinction is his assumption of a hierarchy among components of scientific practice based on their temporal endurance: he stresses how general principles and goals (what I hitherto referred to as 'theoretical perspective') constitute a more fundamental and thus more enduring commitment than laboratory practices or specific models adopted for a limited time to test a specific hypothesis. Such a hierarchy illustrates the different levels of entrenchment that commitments can reach within any researcher's practice: a long-term commitment dictates habits that will not be modified unless there are substantial changes in both the theories and methods used in the field, while a short-term commitment might signal simply the week-long choice to adopt a specific mutant of Arabidopsis to perform a series of experiments – a choice that does not necessarily compromise the long-term results of research, but allows to explore a new terrain and determine whether it might be fruitful. What I wish to add to Galison's remarks is that the endurance of commitments often

[226] Galison actually uses the notion of 'constraint', which he classifies as indicating a 'level of commitment' (1987, 244-250).

depends on the outcomes of research carried out under their guidance. Given this factor, which Galison does not take into account, I disagree with his characterisation of models and theories as necessarily short-term commitments: whether they turn out to be tools granting a transition to better models and/or theories, or whether they become a central dogma for a whole field (as in the case of the transcription-translation model in classical genetics, which can certainly be classified as a long-term commitment[227]), depends on the circumstances of their use as well as the results obtained from them. The same is true of laboratory practices.

8.2.2 *Commitments and Personal Identity*

The previous section contained a number of examples and observations concerning commitments and their relations to epistemic skills. In the next section, I shall highlight the usefulness of considering the link between skills and commitments in the context of my analysis of scientific understanding. Before turning to that, however, I would like to focus more closely on the nature and role of commitments in scientific research. In this section I shall argue that commitments constitute important features of the identity of a researcher: features of which scientists might be conscious or unaware depending on their professional and personal circumstances.

Let me start my analysis from the discussion of commitments proposed by Michael Polanyi in his (1958, 1967). Polanyi conceives of commitments as 'a manner of disposing of oneself' (1967, 302-3): that is, a way of behaving that defines and delimits one's personal identity and way of life. Scientists should be particularly careful about the commitments that they undertake, as, in Polanyi's view, their choice of commitments is strongly related to their beliefs: it constitutes 'a responsible decision, in submission to the compelling claims of what in good conscience I conceive to be true' (1958, 65). Through commitments, scientists gain 'legitimate grounds for the affirmation of personal convictions with universal intent' (1967, 324): Polanyi points to the importance of commitments in shaping one's research practices and conferring credibility (in his view, 'objectivity' – hence the reference to 'universal intents') to one's results.

Polanyi's description does not fit my characterisation in one crucial respect: I do not think that commitments *necessarily* constitute a manifestation of belief or a 'personal conviction' of the person that expresses them (1958, 325). A scientist can, in my view, commit to following a specific course of action or way of thinking without necessarily believing it to be 'true' or helpful towards acquiring true knowledge. Commitments can be undertaken for wholly different reasons than the quest for truth, such as for instance pragmatic requirements (having to exercise a skill that took long to be learnt and that guarantees a prestigious job), aesthetic preferences (committing to buy green equipment in one's lab because one likes the colour, rather than because one believes that the brand fabricating green instruments is of superior quality) or emotions (committing to research on lung cancer because a close relative has died of that illness). Further, commitments can be undertaken both consciously and unconsciously, that is, as a result of a decision-

[227] See Keller (2000, 53) on this point.

making process or as a result of habit or unacknowledged preferences (a point to which I return below). Thus, an individual's beliefs about what constitutes true knowledge might be among the reasons for his or her choice of other commitments; yet, the latter remains primarily a *pragmatic* necessity emerging from the individual's need to act and think in specific ways (most notably, in a skilful way) and towards specific goals.

This modification of Polanyi's view implies a revised account of the putative link between commitments and objectivity. Holding commitments is no guarantee for 'universal intent', as Polanyi puts it. As a researcher, I can commit to following specific courses of action without believing that they will uncover universal truths or that they constitute the best possible way to achieve a certain goal: rather, they could constitute a *convenient* way to achieve that goal – a way that, in the specific context in which I work, reveals itself to be more useful than other, supposedly better, ways. Nevertheless, I believe Polanyi is right in indicating that scientists' commitments are as much linked to their research practice as they are expressions of their personality. Polanyi acknowledges that commitments are strongly influenced by the scientific and social context in which research is conducted. Commitments are in fact both a personal feature (of individual scientists who subscribe to them) and an important factor keeping scientific communities together and defining their epistemic cultures. Each epistemic culture in science has a slightly different set of commitments, depending on the specific methods, beliefs and values that researchers in the group decide to adopt. Indeed, scientific communities might be defined as units of researchers sharing a common set of commitments.

Polanyi's account is based on a deep appreciation for what he calls tacit knowledge, that is, knowledge that researchers possess and use in their everyday practices without being aware of it. In Chapter 2, I criticised Polanyi for shying away from the full implications of such a position: regretfully, he ends up relegating the role of tacit knowledge to a subsidiary position with respect to theoretical knowledge, which, according to Polanyi, remains the main goal and primary concern for scientists. I now wish to critique Polanyi's view on commitments for being equally conservative and not fully building on the implications of recognising tacit (and embodied) knowledge as crucial constituents of scientific practice. In several passages of his work, Polanyi is in fact unclear on whether he views commitments as resulting from a self-conscious choice or whether he thinks of them as stipulated unconsciously - a side effect of the research practices conducted by any one researcher. As I already emphasised, it is important to Polanyi that researchers take responsibility for their commitments – a position that implies the necessity, for scientists, to settle on a commitment only as the result of a decision-making process. At the same time, Polanyi rightly notes that 'personal commitment [..] is involved in all acts of intelligence by which we integrate some things subsidiarily to the centre of our focal attention' (1967, 61). In other words, commitments can also be 'tacit', as they amount to the actions and beliefs uncritically and unconsciously assumed by a researcher in order to pursue his/her central research interests.

I view Polanyi's ambiguity on this point as reflecting a reality that hopefully already emerged from my analysis: indeed, as Polanyi should have explicitly acknowledged, research commitments can be *both* tacit and explicit, motivated and unjustified. This finding fits my approach to commitments as entrenched in the successful accomplishment of specific epistemic activities, such as the ones of modeling and abstracting that I have discussed in Chapter 6. I described commitments as necessary conditions for the acquisition, conceptualisation, interpretation and communication of experiences and data: in short, as crucial tools enabling scientists to pursue their investigations. Some commitments are adopted unconsciously as an apparently obvious aspect of a specific research practice, while other commitments are made explicitly and consciously by scientists who are well aware of their impact on their research. The manner in which a researcher 'disposes of himself' can be the fruit of a conscious decision among several credible alternatives. As we saw when surveying the social skills displayed by successful Arabidopsis researchers, it is very useful for scientists to recognise and prioritise commitments that might favour their career, their ability to collaborate with other scientists and/or to bring funding and popularity to a research topic that they are interested in studying. Many prominent scientific debates are constructed around the clash between different commitments. For instance, the unresolved dissent among evolutionary biologists about the exact patterns of evolutionary change (featuring scientists committed to phyletic gradualism against the advocates of 'punctuated equilibria'); or the age-old debate between catastrophists and uniformitarianism about the rate of mass extinctions.[228]

The self-conscious nature of research commitments is emphasised by sociologist Elihu Gerson (1998) in his analysis of commitments as 'more than a verbal promise', that is, as 'an actual expenditure of resources' routinely required by scientific work (ibid., ch. 2, 2). Gerson underlines the way in which commitments constrain and at the same time enable specific courses of action. He also draws a distinction between obligatory commitments, that is commitments dictated by the nature of the phenomena under scrutiny or the social context in which research is carried out, and non-obligatory commitments, that is choices made by individual researchers given a number of available or foreseeable options (ibid., 3). The distinction between obligatory and non-obligatory commitments seems to me illuminating as it highlights the importance, in certain research contexts, of self-consciously making and maintaining certain commitments *in the face of alternative options*. I actually think that these two types of commitments, as well as the manner in which they are undertaken, are associated to different phases in a scientists' career and research trajectory. At certain junctures in his or her career, such as the start and the end of a PhD or the beginning of a new research project, a researcher is relatively free to choose the instruments and models to use in order to investigate a specific phenomenon, as well as the school of thought and discipline under which to carry out the research. It is in this context of broad choices about communities and methods that non-obligatory commitments play their most important role. Once such broad choices are made, obligatory commitments follow in the form of commitments to the more specific norms, concepts and techniques used within the chosen communities to pursue their goals. Further, both obligatory and non-obligatory commitments can be upheld consciously or unconsciously, depending on the social/material context as well as the personality of the researcher in question. A PhD student might be swayed by mentors towards accepting a postdoctoral job in a specific laboratory without being aware of the theoretical, material

[228] For accounts of these two debates, see Gould (2002) and Glen (1994).

and social commitments involved; another student in the same position might consciously subscribe to the commitments enforced by the same laboratory, or disregard external influences in order to join a group endorsing very different commitments.

While Polanyi's account of commitments remains ambiguous on this crucial issue, Gerson's account of intentionality in research commitments proves too one-sided, insofar as it over-emphasises the self-conscious nature of making commitments. Gerson goes as far as to define commitments as a 'bet that a certain course of action (and not others) will turn out to be the right thing to do' (Gerson 1998, 7). This interpretation stands in square opposition to Polanyi's notion of tacit commitment, as it requires not only that researchers are always aware of their commitments, but also that the stipulation of commitments is always preceded by strategic reasoning about their future consequences. As I already mentioned, I do not share Gerson's belief that research commitments are always the result of strategic thinking. Commitments are not necessarily chosen through conscious deliberation over their prospective utility to a given goal, just as they are not always the expression of what a scientist believes to be true or truth-conducive (as advocated by Polanyi). As exemplified by the case of obligatory commitments, many of the commitments displayed in scientific research are not the result of careful consideration and decision-making, but are rather dictated by the epistemic culture, material environment and/or personality of the researcher who adopts them.

Further, Gerson's notion of 'the right thing to do' is dangerously vague. He could use this definition in a *broad* sense, to imply anything judged as 'right' according to evaluative criteria ranging from scientific ('right' in the sense of scientifically successful) to moral and social ('right' in the sense of ethically valuable or socially useful). However, we could take his definition to imply 'right' in a purely scientific sense: researchers bet on commitments that are expected to grant reliable or innovative results by the standards of the one or more scientific communities in which they work. This *narrow* interpretation is more problematic, as it gives a utilitarian picture of researchers as concerned purely with maximising the scientific utility of their research without regard to other factors, ranging from their feelings and moods to their religious and political beliefs. The definition of commitments as utility-maximising bets excludes commitments made for reasons that have no direct link to one's welfare or professional success, such as one's cultural or moral allegiances; one's feelings, as determined by one's professional as well as private life; and one's preferences, as acquired through past experience and sometimes dictated by nothing more than personal taste. For instance, researchers might be committed to studying Arabidopsis rather than mice because they have moral objections to experimentation on animals; they might be committed to protracting their experiments through the night or not to do it, depending on their private life (for instance, the wish to be close or away from their family in the evening), ability to stay awake and feelings towards working at night (as some people dislike the absence of daylight); or they might be committed to a theoretical perspective out of devotion to and respect for their mentors, who endorsed the same perspective, rather than because they are deeply convinced that this choice will improve the quality or visibility of their research. Thinking of committing as betting on a profitable course of action is like thinking of living as betting on a set of personality traits: just as acquiring a personal identity is as much a matter of choice as it is a matter of circumstances and environment, committing to a set of values and ways of acting is as much the result of a conscious deliberation as it is the result of unconscious habits dictated by the context.[229]

8.2.3 Crucial Conditions for Biological Understanding

Research commitments are at once a basic result and a crucial motivation for the exercise of epistemic skills. As claimed in Chapter 6 through the case of modeling, the exercise of epistemic skills allows researchers to determine which knowledge is relevant to understanding the phenomenon under scrutiny. I can now add that commitments play a crucial complementary role to skills in the individuation of knowledge relevant towards understanding a phenomenon.

Skilful abstracting towards the production and use of either TAIR or NASC models does indeed involve specific commitments, which my analysis in terms of abstracting as an activity helps to analyse and compare. As we have seen in Chapters 5 and 6, the abundance of data about Arabidopsis chromosomes and gene products, as well as the appeal that molecular biology exercises on research sponsors, induced TAIR researchers to adopt the term 'gene' as the central organisational term of GO as well as the starting point for all TAIR modeling efforts. This practical constraint implied that the GO network structures its concepts according to their relation to 'gene', while TAIR visual models are still very far from being able to incorporate information about plant morphology, evolutionary history and higher-level physiology. The initial availability of genomic data need not have shaped later research in this way. GO personnel could have encouraged the development of several types of ontology, each dealing with a different aspect of organismal biology and thus adopting different concepts as key reference points for the categorisation and integration of knowledge. Such alternative ontologies are indeed being elaborated by researchers working on different organisms and in communities distant from the communities in which GO researchers work.[230] Alternatives to GO do exist: TAIR has chosen not to consider them and to pursue their initial choice of GO as the appropriate bio-ontology to use in order to organise and store Arabidopsis data. In other words, the steps involved in the actual realisation of the

---

[229] In this sense, my analysis of commitments is closer to Amartya Sen's work on commitments, freedom and personal identity of human agents (Sen 1976, 96). Sen defines commitments in a way apparently similar to Gerson: 'a person choosing an act that he believes will yield a higher level of personal welfare to him than an alternative that is also available to him' (ibid., 95). Sen clearly states that the choice to commit is geared towards enhancing someone's *personal* welfare. This includes various aspects, such as financial, professional and family welfare, which the person in question can rank depending on his or her preferences and priorities. A person more interested in career than in family life will rank professional welfare as the highest form of personal welfare; a person valuing economic prosperity over social ties will view financial welfare as the highest form of personal welfare. These preferences are in themselves already a broad commitment; further, they are an expression of the individual's personal identity and of its ties to the social context in which the individual finds him or herself. As Sen notes, 'commitment is, of course, closely connected to one's morals. But moral in a very broad sense , covering a variety of influences from religious to political, from the ill-understood to the well-argued' (ibid., 97).

[230] The rapid growth of various types of bio-ontologies is documented in the following website, which collects and gives access to all bio-ontologies including GO: http://obo.sourceforge.net/.

abstracting of Arabidopsis biology into TAIR models generated a theoretical commitment to a gene-centric vision of Arabidopsis biology. This theoretical commitment generated, in turn, a performative commitment to using Java programming to visualise such gene-centric ordering of data; as well as a social commitment to spreading this view (and the resulting tools) among Arabidopsis researchers of all trades. These commitments are explicitly acknowledged by the TAIR team and condemned by many ecologists and evolutionary biologists.

Does NASC modeling commit to gene-centrism as heavily as TAIR does? The answer to this question is, as briefly mentioned in Chapter 7, definitely not. NASC researchers translate their everyday experience with the morphology and physiology of actual plants into a theoretical commitment to highlighting their macroscopic features rather than their microscopic ones. As illustrated in Chapter 6, PATO itself is constructed around concepts descriptively applying to anatomical components of the plants (such as 'leaf" and 'stem'). This theoretical commitment is associated to a performative commitment to descriptive accuracy in categorising ecotypes cultivated within NASC facilities, where researchers have a habit of identifying their plants both through genomic structure and through their morphological traits. Some of the associated social commitments adopted by NASC biologists include the willingness to privilege technical over theoretical work. Rather than valuing discussions and exchanges about theoretical knowledge of Arabidopsis biology, they privilege sharing embodied knowledge of the plants themselves and ways to handle them, thus often seeking the advice of experimental biologists, plant breeders and laboratory technicians. Rather than seeking theoretical results themselves, they are committed to isolating and reproducing ecotypes that will become the best available tools towards acquiring theoretical results.

The focus on the skills required to abstract features of a phenomenon (in order to model it) allows me to trace the commitments involved in abstracting processes at both TAIR and NASC. Once we have traced the commitments implied in the production of each model, it also becomes easier to clarify how those models are handled by the researchers using them and with which epistemological implications. This is true, for instance, in the case of the nagging preoccupation of any biologist engaged in model organism research, which I have raised already in chapter 3: that is, the question concerning the representational value of the results obtained on any specific organism. Broadly, this issue concerns the applicability, or significance, of results obtained through the study of a specific organism. Can those results tell us something about other types of organisms? And how can we determine this? As I already pointed out, within the Arabidopsis community there is a strong drive to interpret results obtained on Arabidopsis specimens as representative not only of all Arabidopsis ecotypes, but of all plants (and even, in some cases, of animal organisms up to Homo sapiens, as in the recent case of the discovery of mitochondrial gene transfers[231]). The extent to which a model organism can be representative of other organisms – that is, the extent to which knowledge of the biology of a specific organism can be taken to be relevant to understanding other types of

[231] Evidence for mitochondrial gene transfers has been one of the most important results hitherto achieved through research on Arabidopsis. A review of this discovery and its impact on biology as a whole is provided by Millar et al (2004).

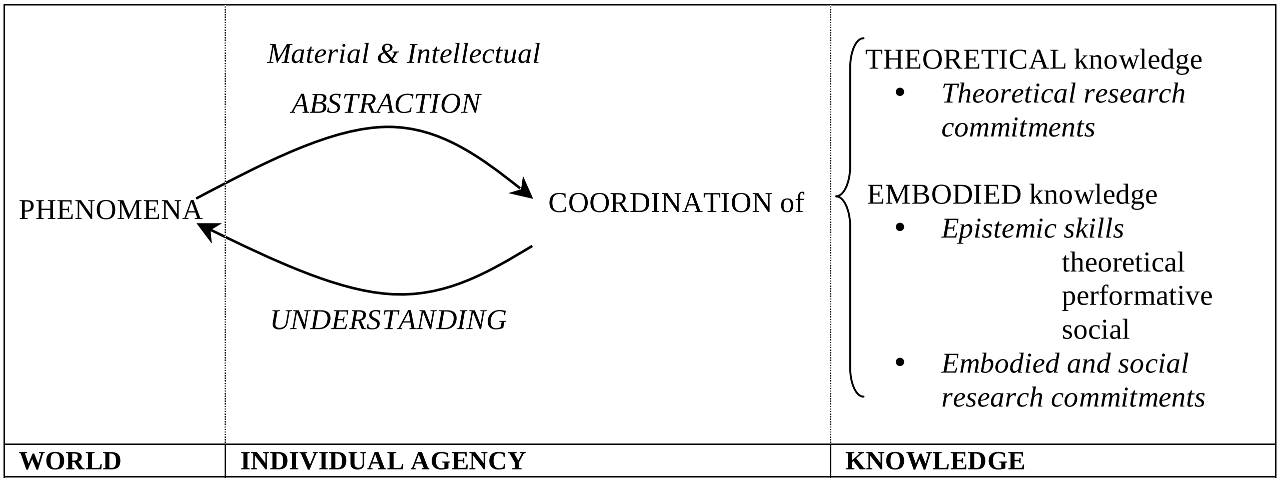
organisms – depends as much on biologists' background knowledge and skills as it depends on their beliefs and objectives. In other words, it depends on the biologists' commitments as much as on their expertise. Acknowledging the skilful practices and relevant commitments involved in abstracting has implications not only for the developers of models but also for their users. Scientists employing NASC specimens and TAIR images to further their understanding of plant biology require skills, materials and commitments that allow them to handle those models *successfully* to this aim.

More broadly, reflection on differences in commitments and skills may also help us to understand the differences in epistemological commitments characterising the experimental versus the theoretical life sciences, as well as the reasons for such differences and – more practically - the possibilities for a better integration of different modeling techniques such as the PATO and the GO terminologies (in the framework of what is sometimes referred to as 'integrative' or 'system' biology; more on this below). Further research is needed on the way in which different communities perceive and adhere to commitments: for instance, as signalled by Galison in his analysis of theoretical versus experimental cultures in physics, 'theorists often divide over the choice of guiding principles while maintaining a consensus on the rules specifying legitimate inferences from them. [..] experimental debates can flourish even when the specific experimental goal is above dispute' (1987, 244).

Together, skills and commitments represent two fundamental tools for the achievement of understanding: adherence to commitments and skilful research practices constitute basic conditions under which researchers coordinate theoretical and embodied knowledge towards understanding specific phenomena.[232]

Table 8.1. Schematic representation of my views on how scientists achieve an understanding of phenomena through coordination of theoretical and embodied knowledge gathered from abstracting relevant features of the phenomena in question.

[232] I hold these two conditions as jointly sufficient, from an epistemological perspective, in order to grant understanding. Yet, I do not intend to portray my account as a complete picture of the conditions that would suffice to guarantee scientific understanding of any biological phenomenon independently of the specific research context. Further, such a complete account would require knowledge of the cognitive processes underlying understanding – knowledge that, despite the fruitful efforts of neurologists and psychologists, is not yet available.



8.3 Kinds of understanding

8.3.1 Theoretical, Embodied and Integrated Understanding

In the conclusion to Chapter 5, I mentioned an important consideration emerging from my view. There can be several different ways of coordinating theoretical and embodied knowledge about a given phenomenon, depending on the actual knowledge available to the researcher seeking understanding of a biological phenomenon. As illustrated by the above comparison between NASC and TAIR modeling activities, researchers possess different combinations of skills and commitments depending on the epistemic culture to which they belong, their goals, training and professional experience. Of course, all researchers, no matter what their field and occupation is, possess a minimal amount of theoretical knowledge as well as theoretical commitments to specific concepts and theories guiding their research; and they possess embodied knowledge, in the form of both theoretical and performative skills, as well as performative commitments. Yet, the balance between theoretical, performative and social skills, as well as commitments, can vary. Depending on such variation, there can be different ways to understand the same biological phenomenon. For instance, the database developed by TAIR can be used to acquire different understandings of Arabidopsis biology, depending on the training and experience of the researchers accessing the resource (e.g. an experimental biologist will use information gathered through TAIR in a different way than a theoretical biologist).

As I tried to make clear, understanding is not a monolithic process. It is not a unique activity to be performed better or worse depending on abilities and resources, nor is it a measurable, homogeneous property applying to a larger or smaller extent to any particular researcher.[233] I shall try to clarify this intuition by examining and comparing three kinds of understanding, each of which derives from different combinations of

[233] The idea of understanding as a ‘property’ that can be measured is often embedded in common-sense uses of the term understanding, as in ‘I understand a little bit of thermodynamics, while Piet understands a lot of it’.

theoretical and embodied knowledge. The first, or *theoretical understanding*, denotes a situation where understanding of biological phenomena is acquired through recourse to theoretical commitments and skills, with performative skills and commitments playing a subsidiary role. The second, or *embodied understanding*, illustrates the opposite situation: theoretical skills and commitments are used to pursue and develop a set of performative commitments and skills, which constitute the main interest of the researcher. The third kind of understanding is the one that interests me the most, as it is the one proposed as a principal aim in Arabidopsis research: that is, the *integrated understanding* deriving from a balanced exercise and coordination of theoretical and embodied knowledge.

Theoretical understanding is based largely on the exercise of theoretical commitments, theoretical knowledge and embodied knowledge (the latter in the form of predominantly theoretical skills). A good illustration for this kind of understanding is provided by population biology, that is, one of the main branches of biology studying evolutionary patterns with the help of mathematical models of populations. In the case of understanding gathered in this field, such as for instance the understanding of the evolving interaction of two populations endowed with different traits, research is accompanied by extensive theoretical commitments: researchers are not questioning what is meant by evolution, how populations and population interactions are characterised, whether traits are inheritable or acquired by nurture, as such assumptions are usually a platform from which they can try to derive results. Similarly, theoretical biologists tend to make strong embodied commitments towards specific ways of looking at phenomena, such as specific diagrammatic or mathematical representations and computer programmes allowing for specific types of simulations (and not others). This is the extent to which such researchers interact with phenomena to be understood: with the help of modeling tools that are conceptually abstracted from the phenomena themselves. Performative commitments and skills are therefore subsidiary to the development of theoretical skills and knowledge: they are not an aim in themselves and embodied knowledge of the phenomena under scrutiny is generally not valued as a crucial component of these researchers' understanding of them.

Theoretical understanding contrasts very strongly with the kind of understanding of biological phenomena obtained, for instance, by a researcher working in the NASC glasshouse. Such a person is specialised in taking care of Arabidopsis specimens and seeds: he or she exercises a number of performative skills, such as sowing, harvesting, cleaning and ordering the seeds, as well as feeding the plants, caring for them and checking on their health and growth. These skills are related to a number of performative commitments, which consist of the standards adopted by NASC for what constitutes a healthy plant and a specific type of seed. Further, most researchers working in the glass-house possess theoretical skills and commitments, such as the skill to fit observed traits into the conceptual framework provided by the PATO system of categorisation and the commitment to produce specimens of the desired ecotype in total isolation, that is, by preventing cross-breeding among different variants. As a consequence of such expertise and daily manoeuvres, NASC researchers acquire an understanding of plants that concerns their macroscopic traits and developmental strategies, rather than their microscopic traits. They understand plant development in its phenomenological expression, a type of understanding that is often not shared by researchers specialising in theoretical analyses of plant development. This is what I call embodied understanding: it revolves around embodied knowledge and makes very few theoretical commitments, as what is central to it is the ability to intervene in phenomena, rather than to explain or predict them. Embodied understanding might in fact be referred to as understanding geared to acquiring control over phenomena. The emphasis of research is on acquiring performative skills that will be useful to explore phenomena through interaction with them. Theoretical skills required to this aim are likely to be minimal: the theoretical knowledge used is the background knowledge needed to formulate questions that might direct explorative interactions with phenomena.

I would like to distinguish the theoretical and embodied types of understanding from a third type of understanding, to whose features and significance I shall devote the rest of this chapter. This is what I call integrated understanding, that is, an understanding resulting from a balanced combination of theoretical and embodied knowledge as well as theoretical and performative commitments. This means that neither theoretical nor embodied knowledge are given a privileged role in the understanding of a phenomenon. Rather, researchers are interested both in acquiring a theoretical interpretation of a phenomenon and in obtaining tools and methods that will enable them to match such theoretical interpretation to the actual features of the phenomenon. To this aim, the embodied knowledge used to acquire this type of understanding includes both theoretical and performative skills: researchers exercise their ability to reason as much as their ability to observe and/or intervene in the phenomena under scrutiny.

8.3.2 Integrative Understanding and Integration in Biology

> *A different model of integration centres on cooperation and communication among theoretical and phenomenal equals, rather than on imperialism and competition for primacy and fundamentality*
>
> *James Griesemer 2006a, 4*

I now wish to propose and discuss the claim that integrated understanding constitutes the best way to understand biological phenomena in a scientific manner. Two objections can be levelled against this proposition, each of which I shall analyse in turn.

First, the quality of understanding acquired by scientists depends largely on their research contexts and goals. For instance, in the context of theoretical biology, researchers need to concentrate all their skills and resources on acquiring theoretical interpretations of phenomena. This means committing to using mathematical modeling, that is, an intellectually abstracted set of models with high explanatory power but little empirical accuracy.[234] Yet, it might be argued, this choice leads to gaining insights that might not have been obtained if biologists insisted on making empirical sense of their findings

[234] I have argued for the need for a trade-off between empirical accuracy and explanatory power in Chapter 6, where I analyse the difference between what I call material and intellectual abstracting in modeling.

throughout their research. In theoretical biology, it seems entirely justified to regard theoretical understanding as the most useful kind of biological understanding. Another example is the field of natural history, where researchers commit to accumulating detailed knowledge about the morphology of as many species as possible. They thus acquire embodied understanding of what organisms look like, where they can be found and how they should be spotted, kept and eventually embalmed. It does not matter that the skills and commitments of natural historians greatly limit their theoretical understanding of the differences among organisms (and, eventually, the causes of such difference): for their research purposes, the best understanding of organisms is embodied rather than theoretical. In view of these examples, how can it be maintained that integrated understanding constitutes a better way to understand biological phenomena regardless of the research field and goals?

The second objection against the preference for integrated understanding looks even more compelling. It could be claimed that it is simply impossible for an individual biologist to acquire integrative understanding, given the increasing specialisation in subfields and the use of specific techniques characterising today's biology. The study of each aspect of an organism, not to mention a population or ecosystem, is conducted with the help of thousands of different techniques, tools and perspectives. The skills and commitments within each epistemic culture are acquired through years of training and require extensive division of labour among biologists, thus making it technically impossible for any one researcher to hold more than a very limited set of skills and theoretical knowledge at the same time. Further, it is not socially advantageous for a researcher to keep shifting from one set of skills to the next: the risk is for his or her work not to be recognised in any of the institutionally established disciplinary structures, thus damaging his or her job prospects, while also losing touch with developments in one field in order to acquire basic skills in another.[235]

I intend to argue that even in the light of these two objections, there are very good reasons for holding integrated understanding to be the most preferable type of understanding in biology. Let me begin from tackling the first issue. It is indeed true that the type of understanding that is privileged within any research community depends on its specific goals and tools. It is also true that there are good reasons for this preference, as each field needs to concentrate its resources on the skills and commitments that best fit its research interests. Yet, this sociological observation does not tell us much about the

[235] As we have seen, such fragmentation has the positive effect of enhancing epistemological diversity in the types of understanding acquired by different biologists of the same phenomena. Epistemic communities are by no means isolated from each other: peer-different collaboration fuels constant interactions and overlaps among them, thus guaranteeing that the boundaries between different epistemic cultures keep shifting and changing depending on the available institutional arrangements and scientific goals. Individual researchers can choose whether to stick to one specific research group for their whole research career or, as is often the case, meander within a few different research communities, thus acquiring skills and commitments helpful to conduct different types of research and acquire (and combine) different kinds of understanding. Yet, because of the tendency to value specialisation over generalist knowledge within biology (which we encountered also in Chapter 7, section 7.3, when discussing TAIR), the number of epistemic communities within which any one researcher can successfully move is limited – so limited, in fact, as to put the sustainability of integrated understanding in question.

overall *quality* of the understanding of biological phenomena obtained by communities focusing exclusively on theoretical or on embodied understanding. The theoretical understanding of evolving systems acquired by theoretical biologists could be viewed as very partial, since these researchers often have no idea of how their results could apply to actual phenomena: they use intellectual abstracting to such an extent that they lose a sense of how their models and theoretical knowledge could be matched to observations, measurements, statistics and other empirical data. Similar problems plague the field of natural history, which has lately lost much of its attractiveness precisely because it does not provide interpretive tools to make sense of the observed differences among organisms and among species.[236] In both examples, researchers seeking *either* theoretical *or* embodied understanding seem to be missing out on something important. After all, biology is about the exploration *as well as* the analysis of phenomena: researchers who understand a phenomenon in one of these two ways should ideally strive to understand it also in the other, thus balancing the amount of theoretical and embodied knowledge used to this aim.

Hence, at least in principle, it makes sense to identify integrated understanding as the kind of understanding that biologists should pursue in the long term, no matter their expertise and specific interests. Now the question becomes, can biologists actually pursue this type of understanding in practice? This brings me to tackle the second objection against integrated understanding. The extent to which an individual can master and combine skills and knowledge as diverse as required to obtain integrated understanding is, as I noted, extremely limited. At the level of the biological community as a whole, there is a sense in which specialisation in biology actually enhances the chance of obtaining an integrated understanding of phenomena. The division of labour in biology, as well as the critical discussions and controversies enabled by such epistemological pluralism, make it possible to acquire skills, commitments and knowledge that facilitate the integrated understanding of at least some biological phenomena and that could not have been acquired without such division of labour. For instance, consider micro arrays technology, which was initially developed by molecular biologists as a tool to store information: researchers in medicine found that they could use micro arrays to measure the susceptibility of genes to environmental conditions, thus enhancing medical understanding of the genetic basis of diseases. In this case, a tool developed thanks to the specific skills and specialised knowledge of one subfield is found to help researchers understand phenomena in another subfield. At the same time, understanding remains, in my account, a cognitive achievement by individuals. This means that, for integrative understanding to be achieved, individuals must exist that can master and apply all the relevant skills, commitments and knowledge accumulated in biology to understanding the phenomenon under scrutiny. This second step, the individual achievement of integrated understanding, is entirely dependent on the collective acquisition of relevant skills and tools[237], yet it represents a different and more difficult stage of research. Integrated

[236] Typically, observations gathered within natural history tend to become interesting only in the light of evolution (thus, through the disciplinary lenses of an evolutionary biologist).
[237] My depiction of the role of a scientific community as a necessary, yet insufficient condition towards understanding a phenomenon in a scientific manner resonates with my account of social skills as necessary to the acquisition of theoretical and performative skills. It is the individual mastery of theoretical and

understanding cannot be distributed among members of a community, but needs to be acquired by an individual trained to integrate and skilfully apply the knowledge collectively accumulated by such community.

Notably, the feasibility of acquiring an integrated understanding of a phenomenon depends strongly on the scale and complexity of the phenomenon itself. If dealing with a relatively simple and well-circumscribed phenomenon, it is within the reach of an individual to acquire the various skills, tools and commitments allowing him or her to understand that phenomenon in an integrated manner. A good example of this is provided by the field of evolutionary developmental biology (or 'evo-devo') This area constitutes a platform for peer-different collaboration between evolutionary, developmental and molecular biologists, leading participants to acquire the various skills required to obtain an integrated understanding of specific aspects of organismic development (most notably, the evolutionary history of epigenetic mechanisms[238]). This kind of understanding is also visible in the Arabidopsis community, where individuals such as Chris Somerville have assembled enough skills, knowledge and experience to understand some components of Arabidopsis biology in an integrated way. Somerville's research focuses broadly on cell biology and more specifically on the role of the cell wall in cellular metabolism. To acquire an integrated understanding of this phenomenon, he studied various related aspects of plant physiology and microbiology, ranging from understanding how cellulose is made to studying the genetic basis for enzymes catalysing polysaccharides. Further, he developed techniques, instruments and databases that can be used to carry out such research, including ways of handling plants and cell samples (Somerville et al, 2004). The skills developed throughout his long career allow him to coordinate his extensive theoretical and embodied knowledge of Arabidopsis cell biology in order to obtain an integrated understanding of the cell wall (its structure, composition, biological significance and relation to other organs).

Attempts to understand highly complex systems (e.g. a whole organism, an evolving population, an ecosystem) in an integrated manner prove to be much more ambitious and questionable than attempts to understand relatively self-contained phenomena such as cellular components or gene functions. In the former case, the skills and commitments needed to understand various aspects of the phenomenon in question are distributed among several research groups: there are very few individuals – possibly no one yet – who can acquire all these skills and commitments at the same time, so as to obtain an integrated understanding of the complex phenomenon under scrutiny. The Arabidopsis community constitutes an excellent case of a collective effort towards acquiring integrative understanding of a complex system. The founders of the Arabidopsis community, as well as many of its current members, have spent much thinking and

---

performative skills that directly determines the quality of the understanding achieved by any one researcher; yet, no individual would master those skills to the extent required for this task, unless he or she was trained within a community of expert peers.

[238] Müller and Newman (2003), for instance, have studied the early developmental history of multicellular organisms with the help of experimental research as well as highly theoretical digital and mathematical models. They thus acquired an integrated understanding of the developmental (epigenetic) mechanisms responsible for originating biological form, particularly in the case of limbs, body symmetries and segmented structures.

efforts towards devising ways to eventually obtain an understanding of Arabidopsis biology as a whole. Arabidopsis research, and particularly projects such as TAIR, thus provide a fitting context in which to examine the value, implications and results of a commitment to integration. It also exemplifies one way in which a large research community can help its members towards acquiring an integrated understanding of some given phenomena: that is, by striving to construct tools that facilitate the coordinated use of the different skills developed within different sections of the community. Since the early 1980s, Arabidopsis scientists have understood the importance of elaborating methods and tools to integrate knowledge.[239] In other words, they have understood that integration in biology, especially in the form of what I call integrated understanding of phenomena, cannot happen purely through the accumulation of knowledge about different aspects of a phenomenon. Rather, it requires the construction of interdisciplinary tools and standards allowing individuals to acquire an integrative understanding without, however, losing the specialised information and perspective provided by their training in each participating sub-discipline.

Let me illustrate this important point by reference to TAIR, a project that represents the pinnacle of scientific efforts towards integration within the Arabidopsis community. TAIR curators summarise their experience by indicating three conditions as essential to obtaining integration among Arabidopsis results:

(a) the formulation of standardised concepts, such as the terms employed in the GO framework;
(b) the provision of material tools for integration, such as the possibility to easily retrieve data about different aspects of Arabidopsis biology (through bioinformatics) and a facilitated access to plant specimens (provided through NASC);
(c) a so-called 'systems approach'[240], which, in the words of Sue Rhee, implies that integration of knowledge about a phenomenon can be achieved only when working *both* on conceptual integrative tools (such as TAIR) *and* applied projects allowing researchers to experimentally manipulate a large variety of aspects of the plant – from the genotypic to the phenotypic.

---

[239] As we have seen, the community has been organised so as to encourage and enhance personal exchanges, peer-different collaboration and use of standardised tools (including the plants themselves). I have mentioned several advantages of such arrangements in previous chapters. Yet another one is the acquired ability to spot phenomena that are not yet being researched, either because no one noticed that no one else was doing it or because no one noticed their existence. According to all Arabidopsis researchers I interviewed, it is impossible to spot such grey areas without having a more systemic view of what goes on in Arabidopsis biology: both Chris Somerville and Sue Rhee argue that the concentration of research efforts on the same organism, joined with the commitments towards integrating knowledge about it, has propelled researchers in directions otherwise inconceivable. Further, they claim that the Arabidopsis research culture is itself changing. Thanks to the current degree of integration, it is now easier for researchers to move from one area of expertise to the next: they can branch out and do things that have not been tried before, rather than specialise in a single area.

[240] I shall not refer here to the various perspectives, put forwards by both scientists and philosophers, of what 'system biology' is or should be. As illustrated by O'Malley and Dupré (2005), this topic is generating some of the hottest and most interesting controversies in contemporary biology: employing my ideas to address the details of this debate would require another dissertation. I shall therefore treat the idea of 'system biology' in its most basic characterisation.

The first two points, which I discussed at length in Chapters 5 and 6, can be rephrased as the conditions under which researchers can hone skills adequate to the integrated understanding of Arabidopsis biology: (a) concerns the intellectual tools needed to develop appropriate theoretical skills, while (b) takes care of the material settings required for biologists to learn how to manipulate models of the plant. Condition (c) is especially interesting for my purposes, as it suggests that integration might emerge only through the balanced coordination of theoretical and embodied knowledge – hence mirroring closely my own definition of how integrated understanding can be obtained. Remarkably, TAIR director Sue Rhee is one of a handful of Arabidopsis researchers who has actually tried to carry out both wide-ranging conceptual work (as Director of TAIR) and experimental work (on the molecular biology of cold acclimation in Arabidopsis, involving experiments on both the genomics and the ecology of the plant) at the same time and with the same degree of involvement. Her motivations for doing that reflect my own motivations for favouring integrated understanding: as she repeatedly stated to me, such a 'system approach' is the only way to effectively achieve integration, especially between studies of the microstructure and developmental mechanisms of the plant and research on its ecology and evolutionary history. Yet, Rhee also complained to me about how hard it was to keep up both expertises: in fact, she did not think that she would be able to pursue both to the same extent any longer.

As I already remarked, the habit of carrying out both conceptual and experimental research is rarely sustainable by an individual within the highly fragmented, specialist cultures characterising biological research. Conforming to the third condition proposed by TAIR curators is thus only possible as a result of the distributed labour (and related understanding) by the whole Arabidopsis community. This resembles the vision proposed by Chris Somerville and his colleagues at the start of research on the weed. Understanding Arabidopsis biology would require recourse to all available tools for the study of a phenomenon, including all the theoretical knowledge accumulated on various aspects of a phenomenon, plus as many of the theoretical and performative skills currently available to the biologists studying it. Further, it involves reaching an awareness of how the many different results acquired on Arabidopsis biology might relate to each other: a form of integration that allows biologists to coordinate the various types of theoretical and embodied knowledge accumulated on the plant to obtain an overall understanding of its biology.

This latter point constitutes, in my view, the greatest challenge to any research community aiming to acquire an integrated understanding of a complex phenomenon. Given the skills and commitments already distributed within the Arabidopsis community, the achievement of an integrated understanding of the plant (as in the goals put forward by MASC) is possible, at least in principle, and might be obtained in the future. However, such understanding is not yet available, precisely because Arabidopsis researchers still have to find ways to integrate these skills, competences, perspectives, and commitments without losing precious information.[241] What individual Arabidopsis researchers need to

[241] Another great obstacle, already mentioned in section 7.3.2, is the increasing gap between over-specialised and integration-oriented research, as exemplified here by the gap between TAIR generalists, or 'super-experts', and Arabidopsis specialists. The challenge for researchers interested in integration is to

pursue, in order to acquire integrated understanding of Arabidopsis biology, is a kind of *integration without unification*: that is, a way to integrate knowledge that is respectful and receptive to the epistemological pluralism embodied by the several fields and competences involved in studying this plant. As I argued in Chapter 5, the TAIR project is far too dependent on a gene-centric perspective to allow for epistemological pluralism: in this sense, it does not fulfil the goals and values that motivated it in the first place. It could be said that TAIR models and databases achieve integration through unification around a specific theoretical framework, that is the GO network of concepts.[242] In so doing, they inevitably dismiss alternative frameworks emphasising other types of data and concepts, such as for instance data coming from ecology and comparative biology.

Whether TAIR could actually be modified (and thus, according to my arguments here, improved) to encompass alternatives to GO remains an open question. TAIR researchers are responsive to critiques and in the future they could, if appropriately challenged by fellow biologists or complementary scientists, try to adopt a more inclusive approach. The problem is that, as the many groups of biologists working with different organisms and bio-ontologies know, there are no clear clues yet as to what such an inclusive approach would look like, or whether it would succeed. The production and use of bio-ontologies for data storage and distribution is still in its infancy. It is simply too early to say whether bioinformatics could provide efficient ways to integrate knowledge, while still granting access to the richness of information and the plurality of perspectives developed in each biological subfield. It is not unreasonable to expect that bioinformatics could, one day, be used to maximise each researcher's chance to understand complex phenomena in an integrated manner. This expectation is, however, unsubstantiated as yet: my analysis of TAIR demonstrates how far biologists are now from achieving integration without eliminating pluralism.

Some philosophers of biology have argued that integration without unification is actually impossible to achieve, as integration will always imply a decrease in epistemological

---

avoid transforming their interests into a specialisation of its own, which would not be accessible to specialists working on restricted aspects of biological phenomena. At the same time, integrative understanding is obtained through the sharing of skills and commitments: as long as many researchers remain interested only in acquiring a very narrow understanding of simple phenomena, they will not be committed to acquiring (the skills necessary to) a more integrated understanding.

[242] This involves a re-assessment of what TAIR, as a service, can actually provide to its users. TAIR curators have been under the illusion (still mentioned in their funding applications) that TAIR could be used as a tool for discovery in biology, that is, it could provide an environment for testing the validity of experimental hypotheses without necessarily having to carry out experiments. According to TAIR curators, much experimental effort goes into discovering and/or testing connections between pathways, gene expression patterns at different loci and other compatible data sets: this effort could be much reduced by reference to TAIR, since the resource allows users to match their hypothesis against the huge amount of data contained in its databases. If this proposal is taken to imply that TAIR should substitute experimental work, I hope to have shown how untenable it is (Chapter 5): a fair amount of performative skills that are only acquired through experimental work are needed to interpret the significance of TAIR results for Arabidopsis biology. TAIR curators should therefore modify this suggestion by pointing to TAIR as a useful complement to experimental research, rather than a replacement for it. This idea is much more attractive, as TAIR could indeed save experimenters a lot of the work that usually goes into digging out relevant evidence and connecting it with results obtained by other researchers - which does not mean that it could, on its own, provide enough information to enable them to make new discoveries.

pluralism in any research programme. One such philosopher, Wim van der Steen, convincingly demonstrates that at least some attempts at integration are bound to be not only useless, but also damaging to research on biological phenomena (1993). He gives the example of the failed attempts to integrate physiological and psychological approaches to the phenomenon of stress in humans. He concludes:

> The field of stress research is actually a collection of loosely interconnected bits of natural history. The point is that stress stimuli, as stimuli, do not have anything in common beyond producing the same kinds of internal states and/or responses. That's all there is to it. It makes no sense to try to develop general laws of stress. There is just a great diversity of phenomena producing the stress response, and we should appreciate diversity when we meet it, and not try to force it into a common mould (1993, 264).

Van der Steen is here primarily concerned about the *theoretical* kind of integration sought by practitioners in stress research. As a conclusion to my discussion of integrated understanding, I want to emphasise two points in contrast with his claims. First, acquiring an integrative understanding need not necessarily include theoretical unification. Rather than trying to modify different theoretical insights to fit them all under the same framework, biologists seeking an integrated understanding might simply wish to learn how to use all of such insights at the same time in order to understand many different aspects of the phenomenon in question at once. In contrast with van der Steen, and in line with arguments put forward by Mitchell (2003) and Rose (1997), I thus maintain that integration in biology does not *necessarily* imply the unification of approaches used by different types of biologists *under the same overall theory or perspective*.[243] The key to the success of a 'system approach' lies in its capacity to include as many perspectives and expertises as possible. Second, the extent to which integrative understanding proves valuable in scientific practice is unavoidably context-dependent. A unified theoretical understanding of stress might prove not only impossible to obtain, but also useless to researchers themselves: a case, as van der Steen notes, of 'pseudo-integration'. However, an integrated understanding of stress, leading for instance to a standardised way to measure it, might be very helpful in a research context such as biomedical research, where the goal is not to obtain a unified theory of stress, but rather to develop better treatments for it, for instance by assessing whether tests and treatments advised by physicians and psychologists are compatible with each other. Depending on the goals and purposes of research, there are cases that demand, in the words of Richard Burian, 'to seek a coherent account of the competing descriptions [of the same phenomenon] provided by the disciplines involved, laden with discipline-specific presuppositions and theoretical commitments' (Burian 1993, 310). Integrative understanding of complex biological phenomena could therefore be held as a regulatory ideal for all branches of biology, while not necessarily constituting an explicit goal for all research projects independently of their contexts. Some branches of research indeed find it useful, in practice, to focus on specialised understandings achieved within single disciplines; other

[243] At least in principle, the coordination of different types of theoretical and embodied knowledge can be achieved without subsuming such knowledge under a unique framework, but rather by learning to use each perspective to gain a different understanding of the specific aspect of the phenomenon under scrutiny.

fields (or groups, as in the case of the Arabidopsis community) thrive, at least in part, thanks to their pursuit of integrative understanding.

8.4 Final Reflections: Understanding Organisms in Complementary Biology

As I pointed out throughout the text, the arguments presented in this thesis bear wide implications for general epistemology, the philosophy and history of biology and biology itself. In concluding, I would like to briefly discuss some of the research paths opened by my analysis. I start by sketching some foreseeable developments in philosophy. I then point to areas where collaboration between philosophers and biologists could encourage developments in the life sciences, thus pursuing the stance on ‘complementary science’ described in Chapter 4. My closing paragraphs are devoted to a specific instance of my work could contribute to current scientific debates: that is, the controversy over the use of ‘virtual organisms’ in experimental biology.

A large area for further philosophical research is related to one of my main claims about the nature of scientific understanding, that is, the *pluralism in understandings* of phenomena that might be acquired by different individuals through scientific research. It is evident from my account that the same biological phenomenon can be understood in a variety of ways, depending on the skills and commitments of the individual(s) involved. The issue then becomes, which type of understanding is best suited for which type of research? In particular, are there types of understanding (and thus specific combinations of skills and commitments) that are more valuable than others in the context of biological research? One way to investigate this crucial evaluative issue could be to construct a detailed taxonomy of scientific understandings and find criteria to establish how each of them fulfils different goals and interests (where these goals and interests can be scientific as well as economic, social or ethical, as for instance when acquiring a scientific understanding of stem cell research for the purpose of evaluating its ethical status).

This could help to develop normative epistemological frameworks allowing to assess the quality of scientific understanding achieved in any given case, as well as, importantly, its relation to truth in science. I have argued that there is no straightforward link between the scientific understanding of a phenomenon and the truth-value of the theories used to understanding that phenomenon. Scientific understanding is the result of the tools, commitments and skills available to scientists at a particular point in time. Such tools are honed through constant negotiation with the material world, and are thus not simply the fruit of ‘cultural trends’ or ‘social settings’ (as radical social constructivists would like to believe). However, precisely because of the specificity of their context and of the motivations and interests guiding their use, these tools and commitments are fallible: they might lead to achieving ‘true’ understanding of the world, but they constitute no guarantee of such ‘true’ understanding. The precise nature of the link between acquiring scientific understanding of a phenomenon and acquiring ‘truthful’ understanding of that same phenomenon is left unclear in my account. What I hope to have offered is a framework to articulate further thoughts about the relation between truth and understanding, with the aim of outlining a conception of scientific understanding that is not entirely dependent on the truth of the knowledge used to understanding, but rather incorporates the subjective perceptions, experiences and skills of the (fallible and gullible, yet determined) individuals attempting to ‘carve nature at its joints’.

Such a normative view could be usefully employed in the context of debates concerning the *role of expertise* in peer-different collaboration among scientists, as well as in public deliberations about scientific results and their implications. For instance, consider my claim that, without adequate scientific training, lay people cannot hope to acquire a scientific understanding of Arabidopsis biology. This does not amount to declaring that they cannot evaluate the social implications of scientific findings in plant biology, nor that they would not acquire any understanding whatsoever from discussing with biologists or reading about such research. In fact, recognising the differences between scientific and other types of understanding allows to acknowledge the existence of inequalities in expertise that would, if ignored, seriously impair the possibility by different sectors of society to discuss and judge the findings and implications of biological research. These claims could be usefully pursued by reference to the excellent work on expertise and public deliberation available within the sociology of science, which I had to largely disregard for the purposes of this thesis.

Further, an evaluative view on the quality of different scientific understandings of phenomena has the potential to inform several ongoing debates in contemporary biology, thus effectively allowing philosophers to act as complementary scientists in the sense proposed in Chapter 4. One of these debates concerns the elaboration of appropriate (institutional and intellectual) conditions for an *integrative study of the biology of organisms*. My critique of TAIR, pointing out the limits imposed by its reliance on a gene-centric perspective, exemplifies a way in which biological research focused on integration can be evaluated and possibly improved through philosophical analysis. Similarly, philosophers could help assessing how new disciplines such as bioinformatics and new approaches such as ‘system biology’ contribute, if at all, to an integrated understanding of organisms (both by scientists and by the general public). A complementary biologist can usefully identify the research commitments held by currently dominant views and propose viable alternatives. This is a way to enhance the variability of the perspectives and commitments informing the production and use of instruments, models and tools used in biological research - such as, in the case of model organism research, bio-ontologies.

A second important debate centres on the role of model organisms in biological research. Given the importance of this debate in the context of my work on Arabidopsis, I wish to conclude by focusing more closely on how my results could inform further reflection on the use of animals, plants and micro-organisms in research. Especially in view of the new technologies recently emerged as a crutch to more traditional experimental biology, such reflection is needed both at the epistemological and at the ethical level. Arguably, as I claimed in Chapter 6, experimental research on actual organisms cannot be substituted with research on virtual models of organisms without substantial changes in the theoretical and embodied knowledge acquired by biologists conducting the investigation. In contexts such as biomedical research, there is little justification for resorting to research methods and tools granting less information about (and possibly a worse understanding of) the phenomena under scrutiny than others. Hence, it could be argued that the proposal, by several prominent scientists, to build 'virtual organisms' as a substitute for real ones in experiments, is misguided. Remarkably, one of the advocates of this position is Chris Somerville himself, who talks about expanding TAIR until it constitutes a 'meta-plant', that is, an ensemble of models and data that can simulate the actual plant in all respects. This virtual plant would not only be serving the same role as actual Arabidopsis plants in laboratories; it would also represent an integration of the knowledge we possess about plants in general. The knowledge gathered around the plant could be used by people working on any other plant as a point of reference for comparisons and the identification of similarities and differences. In short, knowledge about Arabidopsis gathered in the virtual model would become the rule of thumb for assessing diversity among plants (and organisms in general). The virtual plant would then serve as (1) a heuristic tool to find gaps in knowledge; (2) a tool for discovery; (3) and a 'virtual organism' for testing.

This suggestion is fascinating to scientists, scientific sponsors and interested citizens alike, for different reasons - for instance, the promise of cheaper, faster and more efficient ways of conducting research[244] and the possibility to avoid torturing animals or even humans (as in particularly risky clinical trials) – which also implies no need for ethical restrictions on research. I remain, however, highly sceptical of this project for a number of reasons. First of all, the idea that virtual organisms could substitute experiments on actual organisms is preposterous. Manipulating an actual organism provides different kinds of information from the manipulation of a digital one. Virtual organisms can certainly play a heuristic role in experimental biology, thus complementing research on real animals; however, they can hardly be treated as tools for discovery and exploration of phenomena, and much less as a substitute for knowledge gathered through actual experiments. Second, the appealing idea that this procedure would be more convenient and efficient than actual experimentation is also dubious. The production of virtual models of organisms, especially at this level of complication, is extremely expensive, both for the producers of the model and for its users, who have to acquire equipment and skills adequate to the task. Further, running such complex models and simulations is likely to become slower, the more complex the system under examination. Climate modeling constitutes a good example of what happens to models that become extremely complex: it can literally take months to run one programme or query. Finally, the efficiency of these models can be questioned, precisely because of the inherent risk of bringing all available knowledge of an organism under the same theoretical perspective (as in TAIR). This would result in a substantial reduction of the diversity and richness of approaches used to study organisms, with a consequent loss in integrated understanding of their functioning.

This brief discussion of the role of virtual and actual organisms in experimental biology exemplifies some of the ways in which a philosophical analysis of contemporary biological practices may help to address questions that are very much alive in the sciences themselves. Writing in such a 'complementary science' mode also generates

[244] I thank John Duprè for stimulating my thinking in this direction.

potential constraints to my analysis. One of them is that my discussion of understanding in Arabidopsis research is at least partially dependent on current developments and trends in plant biology. Indeed, it is most likely that Arabidopsis will soon cease to have such a central role as model for the whole of plant biology. As knowledge about other plants (tobacco, maize, yeast are just some examples) increases, biological research will start focusing more on the diversity among organisms rather than their common mechanisms and structures, thus increasingly debilitating the credibility of Arabidopsis as representative of so diverse a group of organisms. The future of Arabidopsis as a model organism almost certainly involves shifts to more comparative approaches to plant biology, which might have interesting implications for a philosophical analysis of biological understanding as well as of ways to integrate biological knowledge across different organisms. This said, I do not believe that these shifts will cause my analysis to become invalid or pointless. The recent history of Arabidopsis research constitutes a significant exemplar for modes of research, organisational structures and styles of reasoning that characterise biology at the turn of the 21st millennium. These modes, structures and styles keep changing, as they did throughout the history of science so far. The way in which individual scientists understand the world might change accordingly, and it will be the job of complementary scientists to trace and discuss these new patterns.

Glossary

ABRC	Arabidopsis Biological Resource Centre (American stock centre for Arabidopsis specimens)
AGI	Arabidopsis Genome Initiative (international project that successfully sequenced the Arabidopsis genome)
Annotation	attribution of properties (such as history, expression, function, localisation, keywords, etc.) to an object. In TAIR: the manual selection (by a curator) of data and references relative to a specific gene or pathway to be inserted in the database
AraCyc	visualisation tool for biochemical pathways
AtDB	*Arabidopsis thaliana* database (predecessor of TAIR in the 1990s)
Curator	biologist in charge of annotating biological data into a database (mostly involved both with setting-up relevant representations and with manually annotating each set of data)
DAG	directed acyclic graph used in object-oriented Java software
Developer	IT programmer assisting curators in developing and maintaining the software needed by model organism databases
EBI	European Bioinformatics Institute
FlyBase	international database for Drosophila melanogaster
GMOD	Generic Model system Database
GMOs	genetically modified organisms
GO	Gene Ontology
GOC	Gene Ontology Consortium
MASC	Multinational Arabidopsis Steering Committee, coordinating international research on Arabidopsis
MetaCyc	TAIR tool for the visualisation of information about Arabidopsis metabolism
MIPS	Munich Information Centre for Protein Sequences
NASC	Nottingham Arabidopsis Stock Centre (main stock centre in Europe)
NIH	National Institute for Health (USA)
NSF	National Science Foundation (USA)
Object	what data-types refer to (e.g. genes, markers, sequences, maps)
OBO	Open Biological Ontologies
PATO	Plant Attribute Trait Ontology
PI	Principal Investigator (head of laboratory/project in the life sciences)
POC	Plant Ontology Consortium
PubSearch	TAIR tool developed to search publications relevant to Arabidopsis research
SeqViewer	TAIR tool developed to visualise genomic sequences
TAB	The Arabidopsis Book
TAIR	The Arabidopsis Information Resource
TIGR	The Institute for Genomics Research (Rockville, Maryland)
2010	NSF-sponsored ‘Project 2010’ encompassing all current research on Arabidopsis Functional Genomics

List of Figures

Unless otherwise indicated, figures are courtesy of The Arabidopsis Information Resource.

List of Tables

Bibliography

Achinstein, P. (1983) *The Nature of Explanation*. New York: Oxford University Press.

Allen, J. F. (2001) Bioinformatics and Discovery: Induction Beckons Again. *BioEssays,* 23: 104-107.

Alonso-Blanco, C. and Koornneef, M. (2000) Naturally Occurring Variation in Arabidopsis: an Underexploited Resource for Plant Genetics. *Trends in Plant Science*, 5, 1: 22-29.

Anderson, M.L. (2003) Embodied Cognition: A field guide. *Artificial Intelligence* 149: 91-130.

Ankeny, R.A. (1997) *The Conqueror Worm: An Historical and Philosophical Examination of the Use of the Nematode C. elegans as a Model Organism.* Unpublished Ph.D. Thesis, Department of History and Philosophy of Science, University of Pittsburgh.

Ankeny, R.A. (2000) Fashioning Descriptive Models in Biology: Of Worms and Wiring Diagrams. *Philosophy of Science,* 67, Issue Supplement: S260-S272.

Ankeny, R. A. (2001) Model Organisms as Models: Understanding the 'Lingua Franca' of the Human Genome Project. *Philosophy of Science* 68*,* 3: S251-S261.

Ankeny, R. (2005) Case-Based Reasoning in the Biomedical and Human Sciences: Lessons from Model Organisms. In: Pétr Hajek, Luis Valdés-Villanueva, and Dag Westerståhl (eds.) *Logic, Methodology and Philosophy of Science: Proceedings of the Twelfth International Congress.* London: King's College Press, pp. 229–242.

Ankeny, R.A. (2006) Wormy Logic: Model Organisms as Case-Based Reasoning. In: Creager, A.N.H., Lunbeck, E. and Wise, M. N. *Science Without Laws: Model Systems, Cases, Exemplary Narratives*. Chapel Hill, NC: Duke University Press.

Bailer-Jones, D. (2000) Modelling Extended Extragalactic Radio Sources. *Studies in the History and Philosophy of Modern Physics*, 31, 1: 49-74.

Bailer-Jones, D. (2002) Scientists' Thoughts on Scientific Models. *Perspectives on Science,* 10: 275-301.

Bailer-Jones, D. (2003) When Scientific Models Represent. *International Studies in the Philosophy of Science*, 17: 59-74.

Bard, J. B. L. and Rhee, S. (2004) Ontologies in Biology: Design, Applications and Future Challenges. *Nature Reviews: Genetics*, 5: 213-222.

Barnes, B., Bloor, D. and Henry, J. (1996) *Scientific Knowledge: a Sociological Analysis*. Chicago: Chicago University Press.

Barnes, E. (1992) Explanatory Unification and Scientific Understanding. In: D. Hull, M. Forbes and K. Okruhlik (eds.) *PSA 1992*, 1: 3-12.

Beatty, J. (1997) Why Do Biologists Argue Like They Do? *Philosophy of Science*, 64, Issue Supplement: S432-443.

Beaulieu, A. (2001) Voxels in the Brain. *Social Studies of Science*, 31: 635-680.

Bechtel, W. (1993) Integrating Sciences by Creating New Disciplines: The Case of Cell Biology. *Biology and Philosophy,* 8: 277-299.

Bechtel, W. and Abrahamsen, A. (2005) Explanation: A Mechanistic Alternative. *Studies in History and Philosophy of the Biological and Biomedical Sciences, 36: 421-441*.

Beck, U. (1991) *Politik in der Risikogesellschaft*. Frankfurt/M: Suhrkamp.

Berg, M. and Timmermans, S. (2004) *The Gold Standard*. Temple University Press.

Bijker, W. (1995) *Of Bicycles, Bakelites and Bulbs: Towards a Theory of Sociotechnological Change*. Cambridge, MA: The MIT Press.

Bloor, D., and Barnes, B. (1982) Relativism, Rationality and the Sociology of Knowledge. In: Lukes, S. and Hollis, M. (eds.) *Rationality and Relativism.* Cambridge, MA: The MIT Press.

Bonner, J. T. (2002) *Lives of a Biologist*. Cambridge and London: Harvard University Press.

Bosman, J. (2006) Reporters Find Science Journals Harder to Trust, but Not Easier to Verify. *New York Times*, 13/02/2006, p.1.

Botstein, D. and Fink, G. R. (1988) Yeast: an Experimental Organism for Modern Biology. *Science*, 240, 4858: 1439-1443.

Boumans, M. (2001) Ceteris Paribus' Conditions: Materiality and the Application of Economic Theories. *Journal of Economic Methodology*, 8, 1: 11-26.

Boumans, M.J. and A. Beaulieu (2004). Foreword to 'Objects of Objectivity'. *Social Epistemology*, 18, 2-3: 105-108.

Bourdieu, P. (1975) The Specificity of the Scientific Field and the Social Conditions of the Progress of Reason. *Social Science Information,* 14: 19-47.

Bowker, G. C. and Star, S. L. (1999) *Sorting Things Out. Classification and its Consequences*. Cambridge, MA: The MIT Press.

Brenner, S. (2002) Life Sentences. Ontology Recapitulates Philology. *The Scientist*, 16, 6: 12.

Brent, R. (2000) Genomic Biology. *Cell*, 100: 169-183.

Buchwald, Jed Z., ed. (1995) *Scientific Practice: Theories and Stories of Doing Physics*. Chicago: Chicago University Press.

Burian, R. (1993) Unification and Coherence as Methodological Objectives in the Biological Sciences. *Biology and Philosophy,* 8: 301-318.

Callebaut, W. (2005) Again, What the Philosophy of Biology is Not. *Acta Biotheoretica*. 53, 2: 55-56.

Cartwright, N. (1983) *How the Laws of Physics Lie*. Oxford, UK: Clarendon Press.

Cartwright, N. (1999) *The Dappled World*. Cambridge, UK: Cambridge University Press.

Cat, J. (2001) On Understanding: Maxwell on the Methods of Illustration and Scientific Metaphor. *Studies in the History and Philosophy of Modern Physics*, 32, 3:395-441.

Chadarevian, S. de (1998) Of Worms and Programmes: Caenorhabditis elegans and the Study of Development. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 29, 1: 81-105.

Chadarevian, S. de and Hopwood, N., eds. (2004) *Models. The Third Dimension of Science*. Stanford, CA: Stanford University Press.

Chang, H. (2004) *Inventing Temperature*. Oxford: Oxford University Press.

Chang, H. and Leonelli, S. (2005a) Infrared Metaphysics: the Elusive Ontology of Radiation (Part 1). *Studies in the History and Philosophy of Science: Part A*, 36, 3: 477-508.

Chang, H. and Leonelli, S. (2005b) Infrared Metaphysics: Radiation and Theory-Choice and the Ontology of Radiation (Part 2). *Studies in the History and Philosophy of Science: Part A*, 36, 4: 687-706.

Chang, H. (unpublished) Ontological Principles and the Intelligibility of Epistemic Activities. Plenary lecture delivered at the conference 'Philosophical Perspectives on Scientific Understanding', Vrije Universiteit Amsterdam, 26 August 2005.

Chicurel, M. (2002) Bioinformatics: Bringing it All Together. *Nature*, 419: 751-757.

Chong, S. (2006) Investigations Document Still More Problems for Stem Cell Researchers. *Science*, 311: 754-755. 10 February 2006.

Clarke, A.E., and Fujimura, J.H., eds. (1992) *The Right Tools for the Job. At Work in Twentieth-Century Life Sciences*. Princeton, New Jersey: Princeton University Press.

Cooper, G. (1996) Theoretical Modeling and Biological Laws. *Philosophy of Science*, 63, Issue Supplement: S28-S35.

Craver, C.F. (2001) Role Functions, Mechanisms and Hierarchy. *Philosophy of Science*, 68: 31-55.

Creager, A.N.H. (2002) *The Life of a Virus: Tobacco Mosaic Virus as an Experimental Model, 1930-1965*. Chicago: Chicago University Press.

Creager, A.N.H., Lunbeck, E. and Wise, M.N., eds. (2006) *Science Without Laws: Model Systems, Cases, Exemplary Narratives*. Chapel Hill, NC: Duke University Press.

Cummings, R. (1983) *The Nature of Psychological Explanation*. Cambridge, MA: The MIT Press.

Cushing, J.T. (1994) *Quantum Mechanics. Historical Contingency and the Copenhagen Hegemony.* Chicago: Chicago University Press.

Daston, L., ed. (2000) *Biographies of Scientific Objects*. Chicago and London: Chicago University Press.

Dretske, F. I. (1977) Laws of Nature. *Philosophy of Science*, 44: 248-68.

Dupré, J. (1993) *The Disorder of Things*. Cambridge, UK: Cambridge University Press.

Durkheim, E. [1895] (2003) What Is a Social Fact? In: Delanty, G. and Strydom, P. (eds.) *Philosophies of Social Science. The Classic and Contemporary Readings.* Maidenhead, Philadelphia: Open University Press.

Escobar, A. (1994) *Encountering Development*. Princeton, New Jersey: Princeton University Press.

Escobar, A. (1999) After Nature: Steps to an Anti-essentialist Political Ecology. *Current Anthropology*, 40, 1: 1-30.

Evans, J.A. (2004) *Sharing the Harvest? The Uncertain Fruits of Public/Private Collaboration in Plant Biotechnology*. Unpublished PhD Dissertation. Department of Sociology, Stanford University. http://jevans.spc.uchicago.edu/papers/j_thesis.pdf

Evans, J.A. (2006) Start-ups in Science: Entrepreneurs, Diverse Backing, and Novelty Outside Business. In: Ruef, M. and Lounsbury, M. (eds.) *Research in the Sociology of Organizations.* Amsterdam: Elsevier.

Evans, J.A. (2006a) Industry Collaboration and Theory in Academic Science. Unpublished manuscript.

Feldmann, K.A. and Marks, M.D. (1987) *Agrobacterium*-Mediated Transformation of Germinating Seeds of *Arabidopsis thaliana*: a Non-Tissue Culture Approach. *Mol. Gen. Genet.*, 208: 1.

Fox, R. (1971) *The Caloric Theory of Gases from Lavoisier to Regnault*. Oxford, UK: Oxford University Press.

Fraassen, B. C. van (1980) *The Scientific Image*. Oxford: Claredon Press.

Fraassen, B.C. van (2002) *The Empirical Stance*. New Haven and London: Yale University Press.

Freese, W. and Schubert, D. (2004) Safety Testing and Regulation of Genetically Engineered Foods. *Biotechnology and Genetic Engineering Reviews*, 21: 299-324.

Friedman, M. (1974) Explanation and Scientific Understanding. *Journal of Philosophy*, 71: 5-19.

Friedman, M. (2001) *Dynamics of Reason*. Stanford, California: CSLI Publications.

Friends of the Earth (2006) *Who Benefits from GM Crops? Monsanto and the Corporate-Driven Genetically Modified Revolution*. Report 110. http://www.foei.org/publications/pdfs/gmcrops2006full.pdf

Fuller, S. (2000) *The Governance of Science*. Buckingham, Philadelphia: Open University Press.

Fuller, S. (2000a) *Thomas Kuhn: A Philosophical History for Our Times*. Chicago: Chicago University Press.

Fuller, S. (2002) *Social Epistemology*. Second Edition. Indiana, Bloomington: Indiana University Press.

Fuller, S. (2006) *The New Sociological Imagination*. London: Sage.

Galison, P. (1987) *How Experiments End*. Chicago, Illinois: Chicago University Press.

Galison, P. (1997) *Image and Logic: A Material Culture of Microphysics.* Chicago, Illinois: Chicago University Press.

Galison, P.L. (1999) Trading Zone: Coordinating Action and Belief. In Biagioli, M (ed.) *The Science Studies Reader*. New York and London, Routledge.

Garcia-Hernandez, M. et al (2002) TAIR: A Resource For Integrated Arabidopsis Data. *Funct. Integr. Genomics*, 2: 239-253.

Gattei, S., ed. (2003) Special Issue: The Kuhn Controversy. *Social Epistemology*, 17, 2/3:89-324.

Geison, G.L. and Creager, A.N.H. (1999) Introduction@Research Materials and Model Organisms in the Biological and Biomedical Sciences. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 30, 3: 315-318.

Gene Ontology Consortium (2001) Creating the Gene Ontology Resource: Design and Implementation. *Genome Research*, 11, 8: 1425-1433.

Gene Ontology Report (unpublished document accessed in TAIR archive under permission of Sue Rhee)

Gerson, E. M. (1998) *The American System of Research: Evolutionary Biology, 1890-1950*. Unpublished PhD Dissertation. Chicago, Illinois: University of Chicago.

Giere, R. (1988) *Explaining Science: A Cognitive Approach.* Chicago. Illinois: Chicago University Press.

Giere, R. (1999) Using Models to Represent Reality. In: Lorenzo Magnani, Nancy J. Nersessian and Paul Thagard (eds.) *Model-Based Reasoning in Scientific Discovery*. New York: Kluwer Academic/Plenum Publishers.

Giere, R. (1999a) *Science without Laws*. Chicago, Illinois: Chicago University Press.

Glen, W. (1994) *The Mass Extinction Debates: How Science Works in a Crisis*. Stanford, CA: Stanford University Press.

Golinsky, J. (1998) *Making Natural Knowledge: Constructivism and the History of Science*. Cambridge, UK: Cambridge University Press.

Gooding, D. (1990) *Experiment and the Making of Meaning*. Kluwer Academic Publishers.

Gould, S.J. (2002) *The Structure of Evolutionary Theory*. Cambridge, MA: Harvard University Press.

Grene, M. (1974) *The Understanding of Nature*. Dordrecht: Reidel.

Griesemer, J.R. (1990) Material Models in Biology. *PSA 1990*, 2: 79-93.

Griesemer, J.R. (2000) Development, Culture and the Units of Inheritance. *Philosophy of Science*, 67 (Proceedings): S348-368.

Griesemer, J.R. (2004) Three-Dimensional Models in Philosophical Perspective. In: de Chadarevian and Hopwood (eds.), Op. Cit. Pp.433-442.

Griesemer, J.R. (2006) Genetics from an Evolutionary Process Perspective. In: Neumann-Held, E.M. and Rehmann-Sutter, C. (eds.) *Genes in Development*. Duke University Press, pp. 199-237.

Griesemer, J.R. (2006a) Theoretical Integration, Cooperation, and Theories as Tracking Devices. *Biological Theory* 1, 1: 4-7.

Hacking, I. (1983) *Representing and Intervening.* Cambridge: Cambridge University Press.

Hacking, I. (1992) Style for Historians and Philosophers. *Studies in History and Philosophy of Science,* 23: 1-20.

Haraway, D. (1988) Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies,* 14, 3: 575-599.

Haraway, D. (2003) Cyborgs to Companion Species. In: Ihde, D and Selinger, E. (eds.) *Chasing Technoscience*. Indiana, Bloomington: Indiana University Press. Pp. 58-82.

Harding, S.G. (1986) *The Science Question in Feminism*. New York: Cornell University Press.

Harding, S. (1998) Dysfunctional Universality Claims? Scientific, Epistemological and Political Issues. In: *Is Science Multicultural?* Indiana, Bloomington: Indiana University Press. Pp.165-187.

Hardwig, J. (1985) Epistemic Dependence. *Journal of Philosophy,* 82: 335-349.

Harré, R. (2003) The Materiality of Instruments in a Metaphysics for Experiments. In: Radder, H. (ed.) *The Philosophy of Scientific Experimentation.* Pittsburgh: University of Pittsburgh Press.

Harris et al. (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research,* 32: D258-61.

Haugeland, J. (1988[1978]) The Nature and Plausibility of Cognitivism. In: *Having Thought. Essays in the Metaphysics of Mind*. Cambridge, MA: Harvard University Press. Pp. 9-45.

Helleman, A. (2003) Consumer Fear Cancels European GM Research: Policymakers worry that plant research in Europe will be further outpaced by the United States. *The Scientist*, 17: 52.

Helmreich, S. (2001) *Silicon Second Nature. Culturing Artificial Life in a Digital World*. Berkeley, CA: University of California Press.

Hempel, C.G. (1965) *Aspects of Scientific Explanation*. New York: Free Press.

Hempel, C.G. (1966) *Philosophy of Natural Science*. Englewood Cliffs: Prentice Hall.

Hempel, C.G. (2000) *Selected Philosophical Essays*. Cambridge, UK: Cambridge University Press.

Hesse, M. (1966) *Models and Analogies in Science.* Notre Dame, Indiana: University of Notre Dame Press.

Hesse, M. (1974) *The Structure of Scientific Inference*. London and Basingstoke: Macmillan.

Hesse, M. (1980) *Revolutions and Recontructions in the Philosophy of Science*. Brighton: The Harvester Press Limited.

Huala, E. et al (2001) The Arabidopsis Information Resource (TAIR): a Comprehensive Database and Web-Based Information Retrieval, Analysis and Visualisation System for a Model Plant. *Nucleic Acids Research*, 29: 102-105.

Hull, D.L. (2000) The Professionalisation of Science Studies. Cutting Some Slacks. *Biology and Philosophy*, 15: 61-91.

Hwang, W.S. et al. (2004) Evidence of a Pluripotent Human Embryonic Stem Cell Line Derived from a Cloned Blastocyst. *Science*, 303: 1669-1674.

Jardine, L. (2000) *Ingenious Pursuits*. London: Abacus.

Jasanoff, S. (1990) *The Fifth Branch: Science Advisers as Policymakers.* Cambridge, MA: Harvard University Press.

Kay, L.E. (2000) *Who Wrote the Book of Life? A History of the Genetic Code*. Stanford, CA: Stanford University Press.

Keller, E.F. (1999) Understanding Development. *Biology and Philosophy*, 14: 321-330.

Keller, E.F. (2000) *The Century of the Gene*. Cambridge, MA: Harvard University Press.

Keller, E.F. (2002) *Making Sense of Life.* Cambridge, MA: Harvard University Press.

Kitcher, P. (1981) Explanatory Unification. *Philosophy of Science*, 48: 507-31.

Kitcher, P. and W. Salmon, eds. (1989), *Scientific Explanation*. Minneapolis: University of

Minnesota Press.

Kitcher, P. (1993) *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford: Oxford University Press.

Kleiner, S.A. (1999) Serendipity and Vision: Two Methods for Discovery. Comments to Nickles. *Biology and Philosophy,* 14: 55-64.

Knorr Cetina, K. (1999) *Epistemic Cultures*. Cambridge, MA: Harvard University Press.

Koertge, N. (1998) *A House Built on Sand: Exposing Postmodernist Myths about Science*. New York: Oxford University Press.

Kohler, R.E. (1994) *Lords of the Fly: Drosophila Geneticists and the Experimental Life*. Chicago, Illinois: Chicago University Press.

Konboy, K., Medina, N. and Stanbury, S., eds. (1997) *Writing on the Body. Female Embodiment and Feminist Theory*. New York: Columbia University Press.

Krieger, C.J., Zhang, P., Mueller, L., Wang, A., Paley, S., Arnaud, M., Pick, J., Rhee, S.Y., and Karp, P. (2004) MetaCyc: Recent Enhancements to a Database of Metabolic Pathways and Enzymes in Microorganisms and Plants. *Nucleic Acids Research,* 32: D438-42.

Krimsky, S. (2003) *Science in the Private Interest*. Rowman & Littlefield Publications.

Krogh, A. (1929). Progress in Physiology. *American Journal of Physiology*. 90: 243-251.

Kuhn, T.S. (1970) Reflections on My Critics. In: Lakatos, I. and Musgrave, A. (eds.) *Criticism and the Growth of Knowledge*. London: Cambridge University Press, pp. 231-278.

Kuhn, T.S. (1977) *The Essential Tension*. *Selected Studies in Scientific Tradition and Change.* Chicago, Illinois and London: Chicago University Press. Pp.320-339.

Lacey, H. (1999) *Is Science Value Free? Values and Scientific Understanding*. London and New York: Routledge.

Laibach, F. (1943) Arabidopsis thaliana (L.) Heynh. als Object fur Genetische und Entwicklungsphysiologische Untersuchungen. *Bot. Archiv*. 44: 439-455.

Lafollette, H. and Shanks, N. (1997) *Brute Science. Dilemmas of Animal Experimentation*. London: Routledge.

Lakatos, I. (1970) Methodology of Scientific Research Programmes. In: Lakatos, I. and Musgrave, A. (eds.) *Criticism and the Growth of Knowledge*. Cambridge, UK: Cambridge University Press, pp. 91-196.

Lakoff, G. and Johnson, M. (1999) *Philosophy in the Flesh. The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.

Latour, B (1987) *Science in Action*. Cambridge, MA: Harvard University Press.

Latour, B. ([1984] 1988) *The Pasteurisation of France*. Cambridge, MA: Harvard University Press.

Latour, B. (1988a) The Politics of Explanation. In: Woolgar, S. (ed.) *Knowledge and Reflexivity. New Frontiers in the Sociology of Knowledge*. London: Sage Publications. Pp. 155-177.

Latour, B. and Woolgar, S. (1979) *Laboratory Life: The Social Construction of Scientific Facts*. Los Angeles: Sage Publications.

Laubichler, M. and Müller, G.B., eds. (2006) *Modeling Biology. Structures, Behaviours, Evolution*. Cambridge, MA: The MIT Press.

Laudan, L. (1990) *Science and Relativism: Dialogues on the Philosophy of Science*. Chicago, Illinois: Chicago University Press.

Ledoux, L., Huart, R. and Jacobs, M. (1974) DNA-Mediated Genetic Correction of Thiamineless Arabidopsis thaliana. *Nature,* 249: 17-21.

Lenoir, T. (2004) Operationalising Kant. In: Michael Friedman (ed.) *Kant and the Sciences,* Cambridge, MA: MIT Press.

Leonelli, S. (2004) Introducing the GJSS: Why a Graduate Journal on Interdisciplinary Methodology? *Graduate Journal of Social Science,* 1, 1: i-vii.

Leonelli, S. (2005) Pluralism and Normativity in Interdisciplinary Research. *Graduate Journal of Social Science,* 2, 1: i-vi.

Leonelli, S. and Reydon, T. (2005) Philosophy of Biology in the Netherlands and Flanders. *Acta Biotheoretica,* 53, 2: 55-56.

Leonelli, S. (2006) Cultivando Hierba, Produciendo Conocimiento. Una Historia Epistemologica de Arabidopsis thaliana. In: Suarez, E. (ed.) *Variedad Sin Limites*. *Las Rapresentationes en la Ciencia*. Universida Nacional Autónoma de México y Editorial Limusa.

Levins, R. (1984). The Strategy of Model Building in Population Biology. In Elliott Sober (ed.), *Conceptual Issues in Evolutionary Biology* (First ed., pp. 18–27). Cambridge, MA: MIT Press.

Logan, C.A. (2001) [A]re Norway Rats.. Things?': Diversity Versus Generality in the Use of Albino Rats in Experiments on Development and Sexuality. *Journal of the History of Biology* 34: 287-314.

Logan, C.A. (2002) Before There Were Standards: The Role of Test Animals in the Production of Empirical Generality in Physiology. *Journal of the History of Biology* 35: 329-363.

Longino, H. (1990) *Science as Social Knowledge: Values and Objectivity in Scientific Inquir*y. Princeton, N.J.: Princeton, New Jersey: Princeton University Press.

Longino, H. (2002) *The Fate of Knowledge.* Princeton and Oxford: Princeton, New Jersey: Princeton University Press.

Lynch, M.E. (1988) Sacrifice and the Transformation of the Animal Body into a Scientific Object: Laboratory Culture and Ritual Practice in the Neurosciences. *Social Studies of Science* 18, 2: 265-289.

Machamer, P., Darden, L. and Craver, C. (2000) Thinking about Mechanisms. *Philosophy of Science*. 67: 1-25.

Magnani, L. (2001) *Abduction, Reason and Science. Processes of Discovery and Explanation.* New York: Kluwer/Plenum Publishers.

Maleszka, R., DeCouet, H. G. and Gabor Miklos, G. L. (1998) Data Transferability from Model Organism to Human Beings: Insights from the Functional Genomics of the Flightless Region of Drosophila. *Proceedings of the National Academy of Sciences of the United States of America,* 95, 7: 3731-3736.

The Multinational Arabidopsis Steering Committee (1999) Realising the Potential of Plant Genomics: From Model Systems to the Understanding of Diversity. Report of the NSF-sponsored Workshop: 'New Directions in Plant Biological Research'. www.arabidopsis.org/info/carnegieworkshop.jsp

The Multinational Arabidopsis Steering Committee (2002) Beyond the Whole Genome Sequence. Report. http://www.nsf.gov/pubs/2002/bio0202/executive.htm

The Multinational Coordinated *Arabidopsis thaliana* Functional Genomics Project (2004). Annual Report. http://arabidopsis.info/info/masc_june_04.pdf

Maynard Smith, J. (1979[1975]) *The Theory of Evolution*. Harmondsworth: Penguin Books Ltd.

Mayr, E. and Provine, W., eds. (1982) *The Evolutionary Synthesis*. Cambridge, Massachussets: Harvard University Press.

Mayr, E. (1982) *The Growth of Biological Thought*. Cambridge, MA: Belknap Press.

Mayr, E. (2004) *What Makes Biology Unique? Considerations on the Autonomy of a Scientific Discipline*. Cambridge University Press. New York.

Meinke, D. W. et al. (1998) Arabidopsis thaliana: A Model Plant for Genome Analysis. *Science*, 282: 672-682.

Meinke, D. and Scholl, R. (2003) The Preservation of Plant Genetic Resources. Experiences with Arabidopsis. *Plant Physiology*, 133: 1046-1050.

Mendelsohn, E., ed. (1984) *Transformation and Tradition in the Sciences*. Cambridge, UK: Cambridge University Press.

Mendelsohn, E. and Green, M., eds. (1976) *Topics in the Philosophy of Biology*. Dordrecht: Reidel.

Meyerowitz, E.M. (2001) Prehistory and History of Arabidopsis Research. *Plant Physiology*, 125: 15-19.

Millar, A.H., Day, D.A. and Whelan, J. (2004) Mitochondrial Biogenesis and Function in Arabidopsis. *The Arabidopsis Book*, 48, 1: 1-36.

Mitchell, S. (2003) *Biological Complexity and Integrative Pluralism.* Cambridge, UK: Cambridge University Press.

Morgan, M. (2003) Experiments without Material Intervention. In: Radder, H (ed.) *The Philosophy of Scientific Experimentation*. Pittsburgh: University of Pittsburgh Press.

Morgan, M. (2005) Experiments versus Models: New Phenomena, Inference and Surprise. *Journal of Economic Methodology*, 12, 2: 317-329.

Morgan, M.S. and Morrison, M. (1999) *Models as Mediators.* Cambridge, UK: Cambridge University Press.

Morrison, M. (2002) Models and Statistics: Pearson and Fisher on Mendelian Populations. *British Journal for the Philosophy of Science,* 53: 39-68.

Myers, N. (forthcoming) Animating Mechanisms, Enlivening Models: Intra-Animacy in the Lively Arts of Protein Modeling. *Social Studies of Science*.

Mueller, L., Zhang, P. and Rhee, S.Y. (2003) AraCyc: A Biochemical Pathway Database for Arabidopsis. *Plant Physiology*, 132: 453-460.

Müller, G.B. and Newman, S.A. (eds.) (2003) *Origination of Organismal Form: Beyond the Gene in Developmental and Evolutionary Biology*. Cambridge, MA: MIT Press.

Mulkay, M. (1981) Action and Belief or Scientific Discourse? A possible way of ending intellectual vassalage in social studies of science. *Philosophy of the Social Sciences* 11: 163-72.

Nader, L. (1974) Up the Anthropologists: Perspectives Gained from Studying Up. In: Hymes, D. (ed.) *Reinventing Anthropology*. New York: Vintage. Pp. 284-311.

Nauta, L.W. and de Vries, G. (1979) Maatschappij of Methode: Bijdragen tot het Internalisme/Externalisme-Debat. *Kennis en Methode*, 3, 1:1-200.

Nickles, T. (1986) Remarks on the Use of History as Evidence. *Synthese*, 69: 253-266.

Nickles, T. (1988) Reconstructing Science: discovery and experiment. In: Batens, D. and van Bendegem, J.P. (eds) *Theory and Experiment*. Dordrecht and Boston: Reidel, pp. 33-53.

Nickles, T. (1995) Philosophy of Science and History of Science. *Osiris,* 10: 139-163.

Nickles, T., ed. (2002) *Thomas Kuhn.* Cambridge, UK: Cambridge University Press.

The National Plant Genome Initiative (2002) *Objectives for 2003-2008*. Washington, DC: National Academies Press. http://www.nap.edu/catalog/10562.html#toc

National Science Foundation (2000) *America's Investment in the Future*. Low & Associates Inc. http://www.nsf.gov/about/history/nsf0050/pdf

O'Malley, M.A. and Duprè, J. (2005) Fundamental Issues in Systems Biology. *BioEssays*, 27: 1270-1276.

Odenbaugh, J. (2003) Complex Systems, Trade-Offs, and Theoretical Population Biology: Richard Levin's "Strategy of Model Building in Population Biology" Revisited. *Philosophy of Science*, 70: 1496–1507.

Orzack, S.H. (2005) Discussion: What, If Anything, Is "The Strategy of Model Building in Population Biology?" A Comment on Levins (1966) and Odenbaugh (2003). *Philosophy of Science*, 72: 479-485.

Oudshoorn, N. and T. Pinch, eds. (2003) *How Users Matter: The Co-construction of Users and Technology*. Cambridge, MA: MIT Press. Pp. 67-79.

Oyama, S. (1985) *The Ontogeny of Information: Developmental Systems and Evolution*. Cambridge, UK: Cambridge University Press.

Paracer, S. and Vernon, A. (2000) *Symbiosis. An Introduction to Biological Associations.* Oxford, UK: Oxford University Press.

Pennisi, E. (2000) Stalking the Wild Mustard. *Science,* 290, 5499: 2055-2057.

Pigliucci, M. (1998) Ecological and Evolutionary Genetics of Arabidopsis. *Trends in Plant Science*, 3, 12: 485-491.

Pigliucci, M. (2003) From Molecules to Phenotypes? The Promise and Limits of Integrative Biology. *Basic and Applied Ecology*, 4: 297-306.

Polanyi, M. (1958 [1998]) *Personal Knowledge. Towards a Post-Critical Philosophy*. London: Routledge.

Polanyi, M. (1967) *The Tacit Dimension*. London: Routledge.

Puccia, C.J. and Levins, R. (1985) Qualitative Models. In: *Qualitative Modeling of Complex Systems. An Introduction to Loop Analysis and Time Averaging*. Cambridge, MA: Harvard University Press. Pp. 1-11.

Putnam, H. (1962) What Theories Are Not. In: Nagel, E., Suppes, P. and Tarski, A. (eds.) *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress.* Stanford, CA: Stanford University Press. Pp.240-251.

Rabinow, P. (1997) *Essays in the Anthropology of Reason*. Princeton, New Jersey: Princeton University Press.

Radder, H. (1996) *In and About the World. Philosophical Studies of Science and Technology*. New York, Albany: State University of New York Press.

Radder, H. (2002) How Concepts Both Structure the World and Abstract from It. *Review of Metaphysics*, 52: 581-613.

Radder, H., ed. (2003) *The Philosophy of Scientific Experimentation*. Pittsburgh: Pittsburgh University Press.

Radder, H. (2003a) Technology and Theory in Experimental Science. In: Radder, H. (ed.) *The Philosophy of Scientific Experimentation.* Pittsburgh: Pittsburgh University Press. Pp. 1-18.

Radder, H. (2006) *The World Observed/The World Conceived*. Pittsburgh: Pittsburgh University Press.

Rader, K.A. (2004) *Making Mice: Standardising Animals for American Biomedical Research, 1900-1955*. Princeton: Princeton, New Jersey: Princeton University Press.

Redei, G.P. (1970) Arabidopsis thaliana (L.) Heynth. A Review of the Genetics and Biology. *Bibliogr. Genet*. 20: 1-151.

Reiser, L., Mueller, L.A. and Rhee, S.Y. (2002) Surviving in a Sea of Data: A Survey of Plant Genome resources and Issues in Building Data Management Systems. *Plant Molecular Biology* 48: 59-74.

Regt, H.W. de (1999) Ludwig Boltzmann's Bildtheorie and Scientific Understanding. *Synthese*, 119: 113-134.

Regt, H.W. de (2001) Spacetime Visualisation and the Intelligibility of Physical Theories. *Studies in the History of Modern Physics*, 32, 2: 243-265.

Regt, H.W. de (2004) Discussion Note: Making Sense of Understanding. *Philosophy of Science* 71: 98-109.

Regt, H.W. de and Dieks, D. (2005) A Contextual Approach to Scientific Understanding. *Synthese*. 144: 137-170.

Rhee, S.Y. et al (2000) Bioinformatic Resources, Challenges and Opportunities Using Arabidopsis as a Model Organism in a Post-Genomic Era. *Plant Physiology*, 124: 1460-1464.

Rhee, S.Y. et al (2003) The Arabidopsis Information Resource (TAIR): a Model Organism Database Providing a Centralised, Curated Gateway to Arabidopsis Biology, Research Materials and Community. *Nucleic Acid Research*, 31, 1: 224-228.

Rhee, S.Y. (2004) Carpe Diem. Retooling the 'Publish or Perish' Model into the 'Share and Survive' Model. *Plant Physiology*, 134: 543-547.

Rheinberger, H. J. (1997) *Toward a History of Epistemic Things*. Stanford, CA: Stanford University Press.

Rich, B. (1994) *Mortgaging the Earth: The World Bank, Environmental Empoverishment and the Crisis of Development*. London: Earthscan Publications Ltd.

Roebbelen, G. (1964) Editorial. *AIS newsletter*, 1:1.

Rose, S. (1997) *Lifelines.* Oxford, UK: Oxford University Press.

Rosenberg, A. (1994) *Instrumental Biology, or, The Disunity of Science*. Chicago, Illinois: The Chicago University Press.

Rosenberg, A. (2001) How Is Biological Explanation Possible? *British Journal of Philosophy of Science*, 52: 735-760,

Rouse, J. (1987) *Knowledge and Power: Toward a Political Philosophy of Science*. Ithaca, N.Y.: Cornell University Press.

Rowan, A.N. (1984) *Of Mice, Models, & Men. A Critical Evaluation of Animal Research*. Albany: State University of New York Press.

Rubin, G.M. (1988) Drosophila melanogaster as an Experimental Organism. *Science,* 240, 4858: 1453-1459.

Ryle, G. (1949) *The Concept of Mind*. Chicago, Illinois: The Chicago University Press.

Saint-Exupery, A. de (2000[1943]) *The Little Prince*. Fort Washington, PA: Harvest Books.

Salmon, W. (1990) *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press.

Salmon, W.C. (1998) *Causality and Explanation*. Oxford, UK: Oxford University Press.

Sanda, S.L. and Amasino, R.M. (1995) Genetic and Physiological analysis of flowering time in the C24 line of Arabidopsis thaliana. *Weeds World*, 2, iii: 1-7.

Schaffner, K.F. (2000) Behaviour at the Organismal and Molecular Levels: The Case of C. elegans. *Philosophy of Science*, 67, Issue Supplement: S273-S288.

Schank, J. and Könhle (2006) Modeling Complex Bio-Behavioural Systems. In: Laubichler, M. and Müller, G.B. (eds.) *Modeling Biology. Structures, Behaviours, Evolution*. Cambridge, MA: The MIT Press.

Scholl, R.L., May, S.T. and Ware, D.H. (2000) Seed and Molecular Resources for Arabidopsis. *Plant Physiology*, 124: 1477-1480.

Sen, A.K. (1976) Rational Fools: A Critique of the Behavioural Foundations of Economic Theory. *Philosophy and Public Affairs*, 6: 317-44.

Shaller, J.E. and Kieber, J.J. (2002) Ethylene. *The Arabidopsis Book*, 12, 1: 1-18. http://www.bioone.org/perlserv/?request=get-document&issn=1543-8120&volume=12&issue=1&page=1

Shapin, S. and Schaffer, S. (1989) *Leviathan and the Air-Pump*. Princeton, New Jersey: Princeton University Press.

Shiva, V. (1999) *Stolen Harvest: The Hijacking of the Global Food Supply*. London: South End Press.

Shiva, V. (2002) *Sustainable Agriculture and Food Security. The Impact of Globalisation*. London: Sage.

Smith, B. (2005) Beyond Concepts: Ontology as Reality Representation. In: Varzi, A. and Vieu, L. (eds.) *Proceedings of FOIS 2004. International Conference on Formal Ontology and Information Systems*.

Solomon, M. (2001) *Social Empiricism*. Cambridge, MA: The MIT Press.

Somerville, C., Flanders, D. and Cherry, M.J. (1997) Plant Biology in the Post-Gutemberg Era. Everything You Wanted to Know and More on the World Wide Web. *Plant Physiology*, 113: 1015-1022.

Somerville, C. and Dangl, J. (2000) Plant Biology in 2010. *Science*, 290, 5499: 2077-2078.

Somerville, C. and Koornneef, M. (2002) A Fortunate Choice: the History of Arabidopsis as a Model Plant. *Nature Reviews: Genetics*, 3: 883-889.

Somerville C., Bauer S., Brininstool G., Facette M., Hamann T., Milne J., Osborne E., Paredez A., Persson S., Raab T., Vorwerk S., Youngs H. (2004) Toward a Systems Approach to Understanding Plant Cell Walls. *Science*, 306. 5705: 2206-11.

Star, S.L. and Griesemer, J.R. (1989) Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, 19: 387-420.

Steen, W.J. van der and Kamminga, H. (1991) Laws and Natural History in Biology. *British Journal for Philosophy of Science*, 42: 445-467.

Steen, W.J. van der (1993) Towards Disciplinary Disintegration in Biology. *Biology and Philosophy*, 8: 259-275.

Steen, W.J. van der (1995) *Facts, Values and Methodology. A New Approach to Ethics. Value Inquiry Book Series*. Amsterdam and Atlanta: Rodopi Press.

Weems, D., Miller, N., Garcia-Hernandez, M., Huala, E. and Rhee, S.Y. (2004) Design, Implementation and Maintenance of a Model Organism Database for *Arabidopsis thaliana*. *Comparative and Functional Genomics*, 5: 362-369.

Suarez, E. (2001) Satellite-DNA: A Case-Study for the Evolution of Experimental Techniques. *Studies in the History and Philosophy of the Biological and Biomedical Sciences*. 32, 1: 31-57.

Suarez, E., ed. (2006) *Variedad Sin Limites. Las Rapresentationes en la Ciencia*. Mexico City: Universida Nacional Autónoma de México y Editorial Limusa.

Suarez, M. (1999) The Role of Models in the Application of Scientific Theories: Epistemological Implications. In: Morgan, M. and Morrison, M. (eds.) *Models as Mediators*. Cambridge, UK: Cambridge University Press. Pp. 168-195.

Suppe, F. (1977) *The Structure of Scientific Theories*. Urbana [etc.], London: University of Illinois Press.

Suppes, P. (1969) A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences. In: *Studies in the Methodology and Foundations of Science. Selected Papers from 1951 to 1969*. Dordrecht: Reidel 1969. Pp. 10-23.

Suppes, P. (2002) *Representation and Invariance of Scientific Structures*. Stanford, CA: CSLI Publications.

Tauber, Alfred I., ed. (1991) *Organism and the Origins of Self*. Edited by R. S. Cohen. Vol. 129, *Boston Studies in the Philosophy of Science*. Dordrecht, Boston, London: Kluwer Academic Publishers.

Thagard, A. (1997) Collaborative Knowledge. *Noûs*, 31, 2: 242-261.

Thomashow, M.F. (2001) So What's New in the Field of Plant Cold Acclimation? Lots! *Plant Physiology*, 125: 89-93.

Toulmin, S. (1961) *Foresight and Understanding: An Enquiry Into the Aims of Science*. New York: Harper Torchbooks.

Trout, J.D. (2002) Scientific Understanding and the Sense of Understanding. *Philosophy of Science*, 69: 212-233.

Trout, J.D. (2005) Paying the Price for a Theory of Explanation: de Regt’s Discussion of Trout (2002). *Philosophy of Science,* 72: 198-208.

Weinberg, S. (1994) *Dreams of a Final Theory*. New York: Vintage.

Wilkinson, M.D. and Links, M. (2002) BioMOBY: An Open-Source Biological Web Services Proposal. *Briefings In Bioinformatics*, 3, 4:331-341.

Wimsatt, W. (1972) Complexity and Organisation. In: Schaffner, K. and Cohen, R. (eds.) *Studies in the Philosophy of Sciences, vol. 20.* Dordrecht: Reidel. Pp. 67-86.

Wimsatt, W.C. (1981) Robustness, Reliability and Overdetermination. M. Brewer and B. Collins (eds) *Scientific Enquiry and the Social Sciences*. San Francisco: Jossey-Bass. Pp. 124-163.

Wimsatt, W.C., (1987) False Models as Means to Truer Theories. In: M. Nitecki and A. Hoffman, eds., *Neutral Models in Biology.* London: Oxford University Press. Pp. 23-55.

Winther, R.G. (2003) *Formal and Compositional Biology as Two Kinds of Biological Theorising*. Unpublished PhD dissertation. Department of History and Philosophy of Science. Indiana University, Bloomington.

Winther, R.G. (2005) Fisherian and Wrightian Perspectives in Evolutionary Genetics and Model-Mediated Imposition of Theoretical Assumptions. *Journal of Theoretical Biology,* 240, 2: 218-232.

Winther, R.G. (2005a) An Obstacle to Unification in Biological Social Science: Formal and Compositional Styles of Science. *Graduate Journal of Social Sciences*, 2, 2: 40-100.

Winther, R.G. (2006) Parts and Theories in Compositional Biology. *Biology and Philosophy*, 21, 4: 471-499.

Woodward, J. (2003) *Making Things Happen*. Oxford, MA: Oxford University Press.

Worrall, J. (1990) Rationality, Sociology and the Symmetry Thesis. *International Studies in the Philosophy of Science*, 4, 3: 305-319.

Wouters, A. (2005) The Function Debate in Philosophy. *Acta Biotheoretica*, 53, 2: 123-151.

Yi, S.W. (2002) The Nature of Model-Based Understanding in Condensed Matter Physics. In: Magnani, L. and Nersessian, N.J., eds. *Mind and Society* (Special Issue: Scientific Discovery, Model-based Reasoning), 5, 3: 81-91.

Zwaan, B. van der and Petersen, A., eds. (2003) *Sharing the Planet: Population - Consumption - Species. Science and Ethics for a Sustainable and Equitable World*. Delft: Eburon Publishers.

Online Resources

ARBC
http://www.biosci.ohio-state.edu/~plantbio/Facilities/abrc/abrchome.htm

Biocurators Archives (for curators of bio-ontologies and model organism databases)
www.biocurator.org

Gene Ontologies
http://www.geneontology.org/

Meyerowitz Lab
http://www.its.caltech.edu/~plantlab/

NASC
http://www.arabidopsis.info

OBOs
http://obo.sourceforge.net/

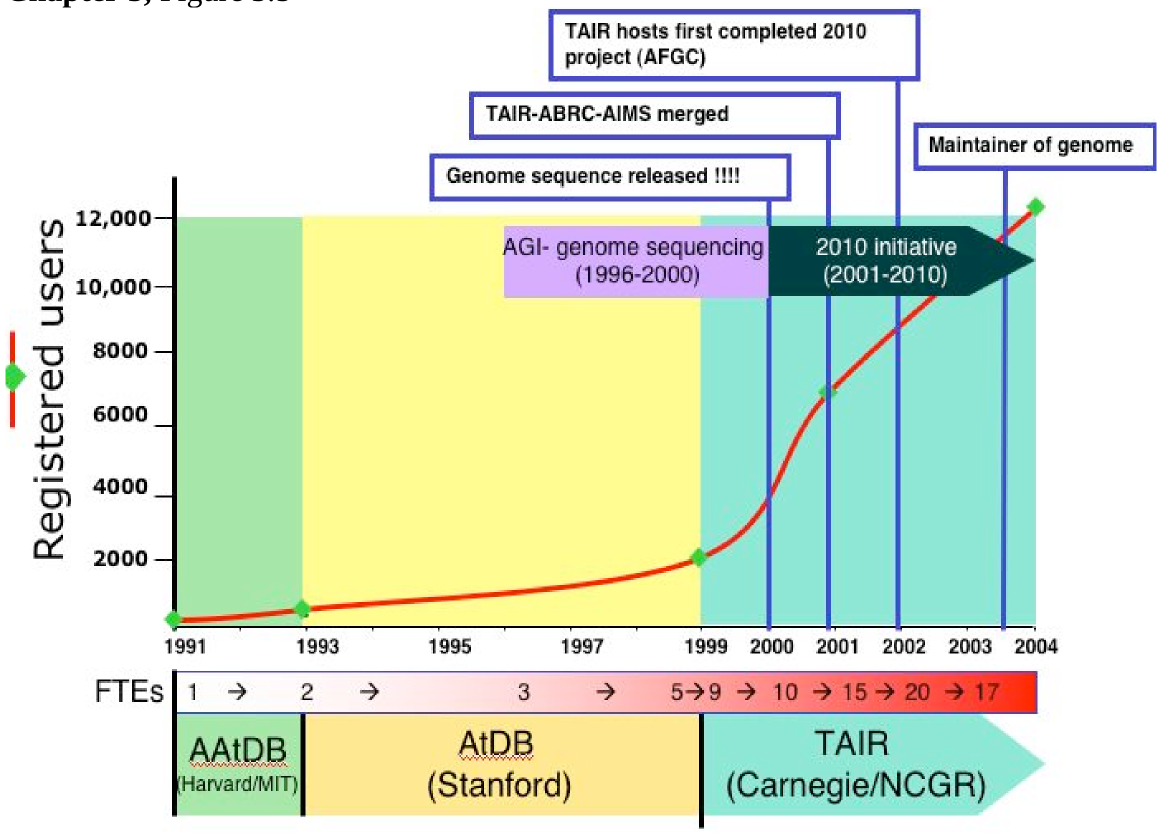
Sue Rhee webpage
http://carnegiedpb.stanford.edu/research/research_rhee.php

TAB
http://www.aspb.org/publications/arabidopsis/

TAIR
www.arabidopsis.org

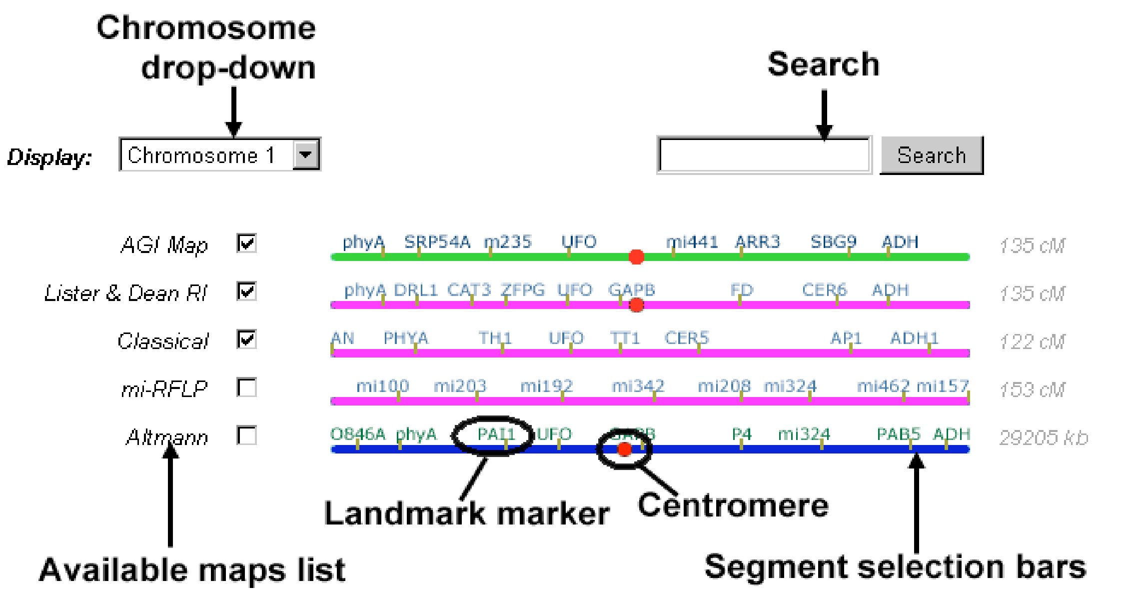
Colour Illustrations

Chapter 3, Figure 3.5

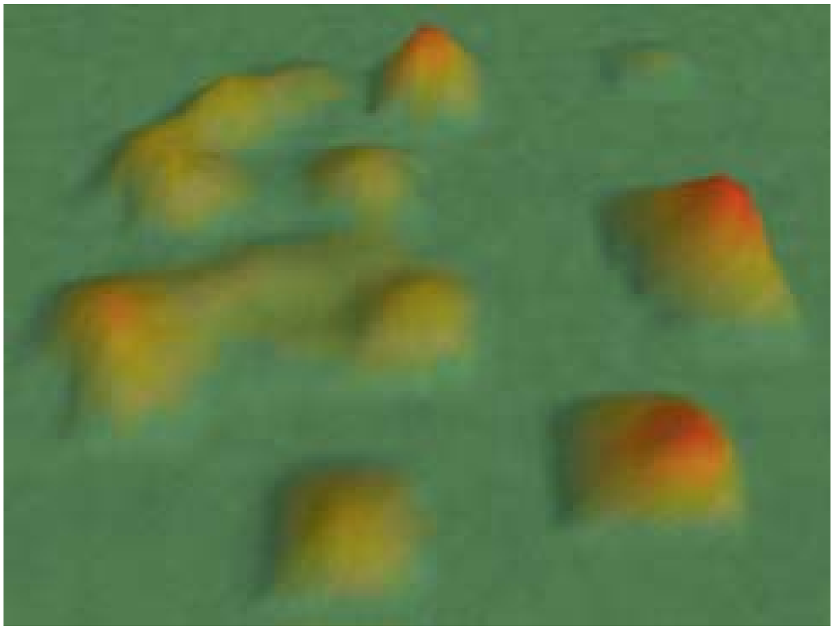


A short history of Arabidopsis Databases

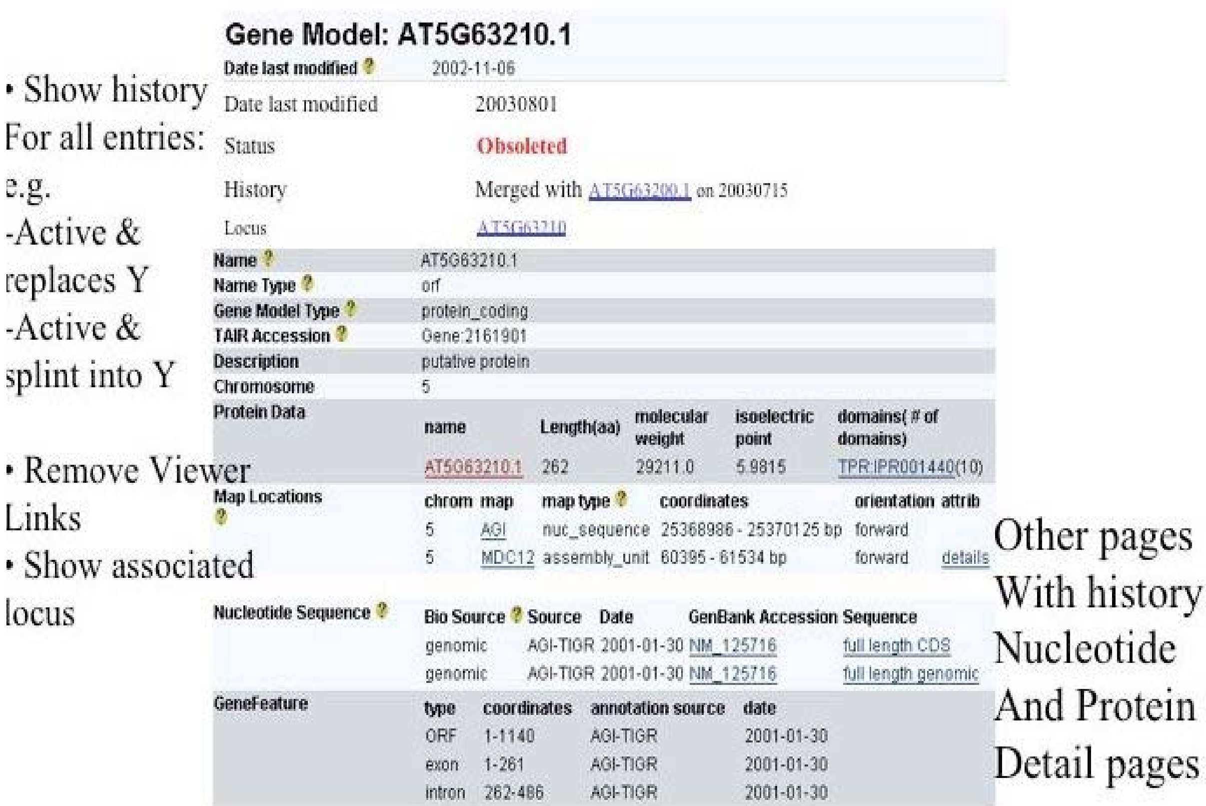
Chapter 5, Figure 5.3



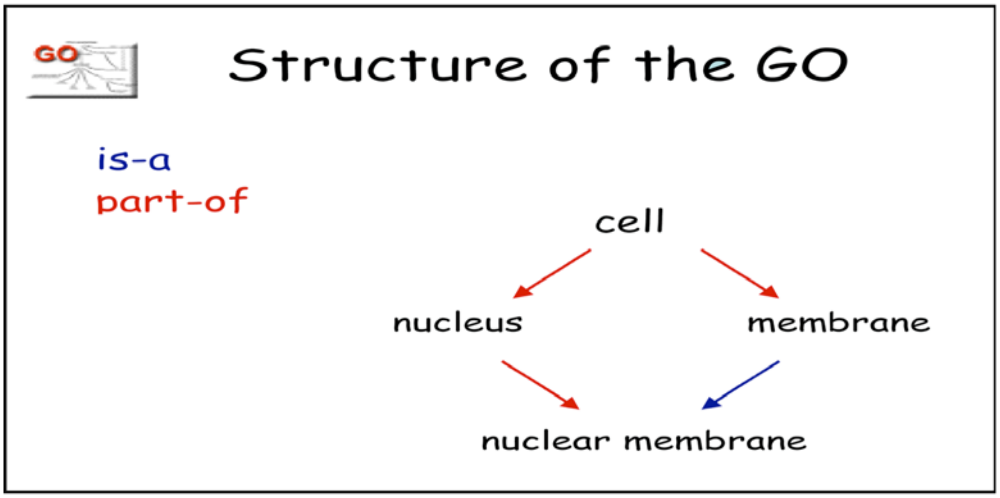
Chapter 5, Figure 5.4



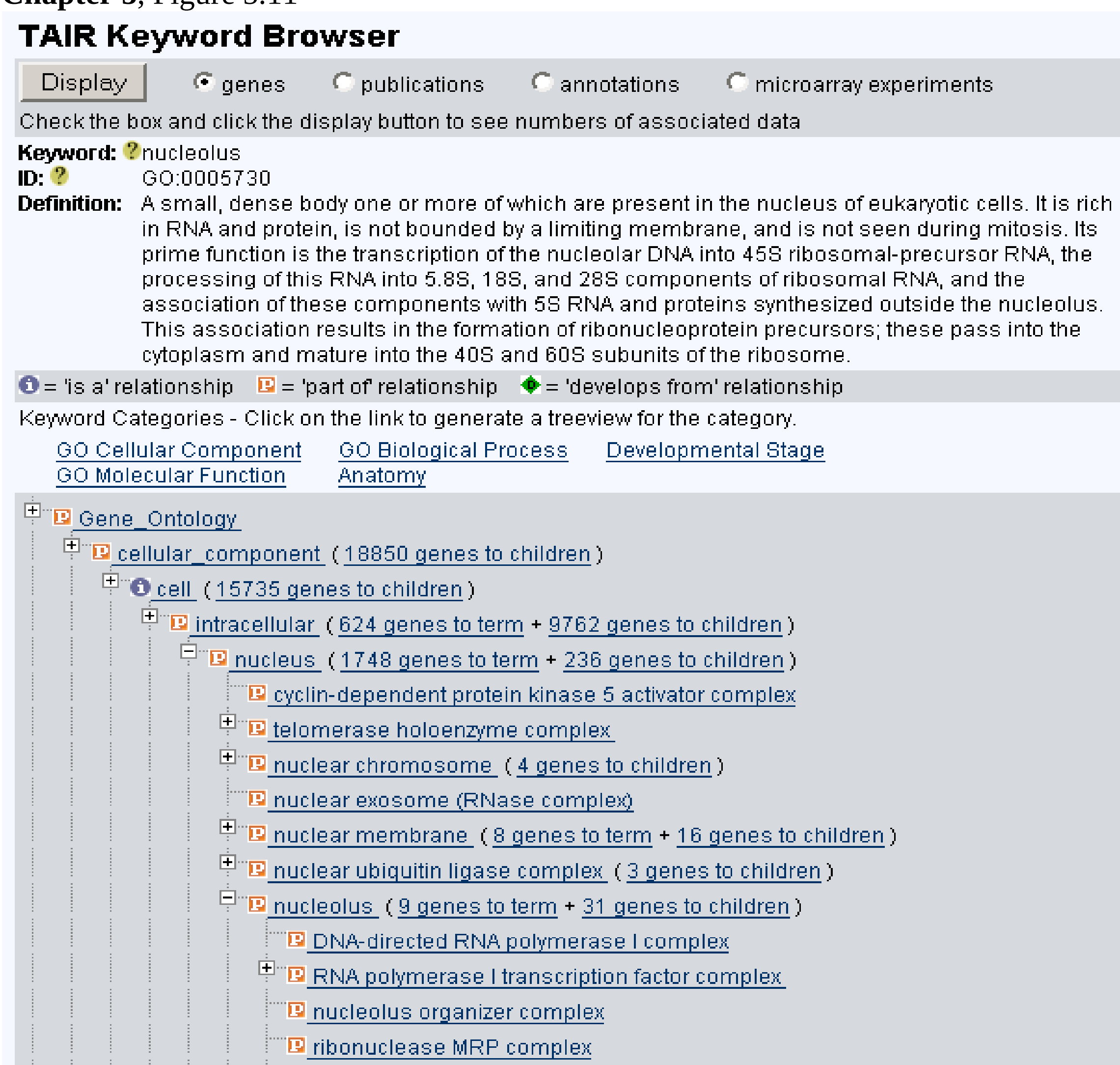
Chapter 5, Figure 5.7



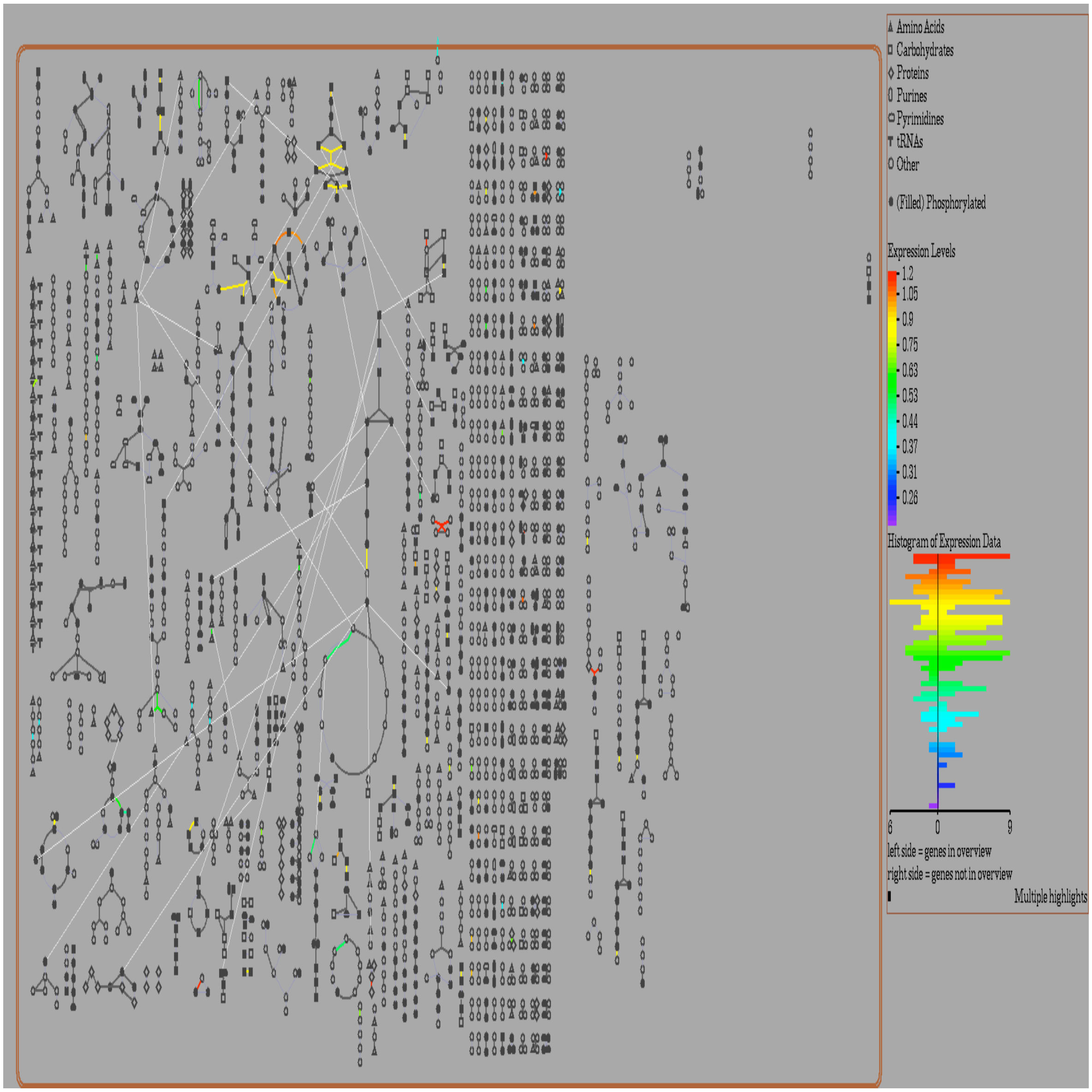
Chapter 5, Figure 5.9



Chapter 5, Figure 5.11



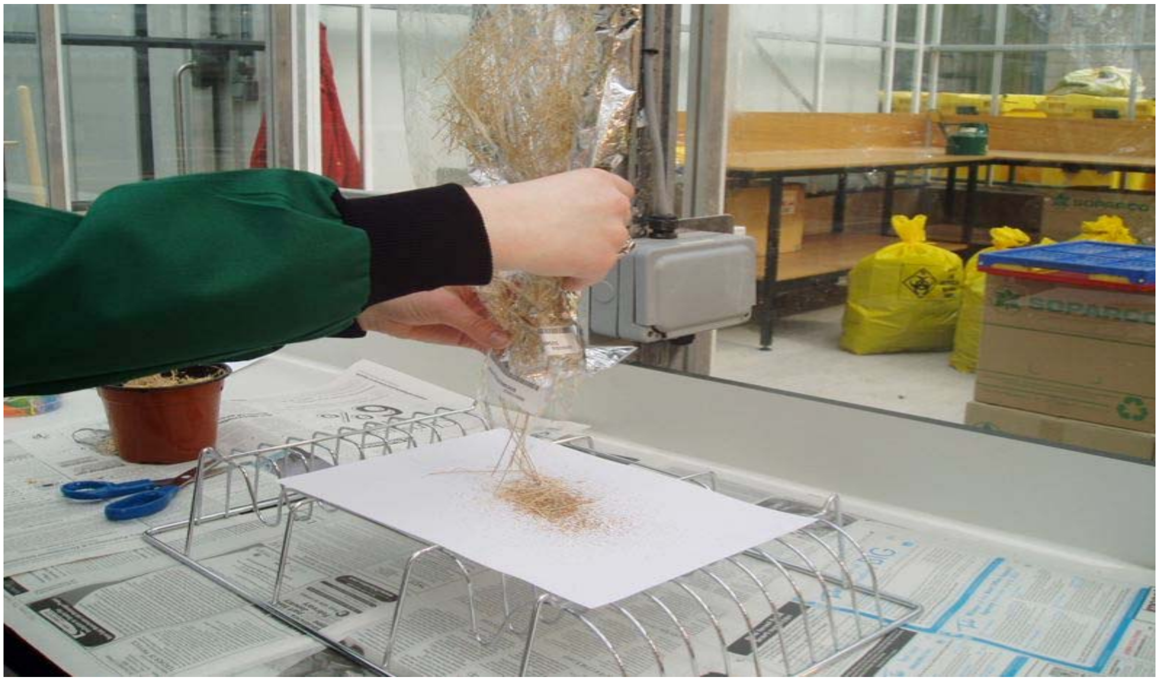
Chapter 5, Figure 5.12



Chapter 6, Figure 6.2

Chapter 6, Figure 6.3

Chapter 6, Figure 6.4



Chapter 6, Figure 6.5

Chapter 6, Figure 6.6

Chapter 6, Figure 6.7

Hoe Biologen Tot Begrijpen Komen:
Onderzoek aan het Modelorganisme *Arabidopsis thaliana*

Nederlandse Samenvatting

Dit proefschrift bestaat uit een epistemologische analyse van noodzakelijke voorwaarden om fenomenen wetenschappelijk te begrijpen, waarbij ik de nadruk leg op de levenswetenschappen. Twee begrippen, al te vaak onder één noemer geschaard in de wetenschapsfilosofie, worden onder handen genomen en hun onderlinge verschillen worden belicht: begrijpen en verklaren. Onder welke voorwaarden leren mensen wetenschappelijke verklaren te gebruiken, teneinde een natuurlijk fenomeen te begrijpen? Wat speelt er bij wetenschappelijk inzicht en het verwerven ervan?

Mijn vertrekpunt bij het behandelen van deze vragen is een nauwgezette historische en sociologische studie van de wetenschappelijke praktijk die het onderzoek aan *Arabidopsis thaliana* (zandraket) kenmerkt, een duinplantje dat het geschopt heeft tot prominent modelorganisme binnen de biologie. Ik bestudeer de procedures, modellen, theorieën en de infrastructuur die specifieke groepen Arabidopsisonderzoekers gebruiken om biologische vraagstukken op te lossen. De relevantie van Arabidopsis als modelorganisme stoelt voornamelijk op het gebruik ervan in interdisciplinair onderzoek in vele takken van de levenswetenschappen. Mijn reconstructie van de geschiedenis en sociologie van het onderzoek aan Arabidopsis dient als een empirische voor een filosofische bezinning over de wijze waarop deze plant wordt gebruikt door wetenschappers om biologische fenomenen op te helderen. Activiteiten zoals het manipuleren van verscheidene fysieke zowel als conceptuele modellen van Arabidopsis zijn cruciaal om de kennis te doorgronden die ontleend is aan de studie van deze plant. Zulke modelleeractiviteiten vereisen verschillende specifieke epistemische vaardigheden en ‘research commitments’, afhankelijk van de materiële eigenschappen van de modellen in kwestie en van hun representatieve functie binnen een gegeven onderzoekscontext. Ieder individu in de Arabidopsisgemeenschap bezit vermoedelijk een verschillende combinatie van relevante vaardigheden en ‘commitments’, en ik argumenteer dat dit de kwaliteit van hun begrip bepaalt.

Ik richt de aandacht op het geheel van vaardigheden en ‘commitments’ die Arabidopsiswetenschappers verwerven door drie types van ervaring: *intellectuele*, waaronder het formuleren van theorieën en verklaringen voor natuurlijke fenomenen en het redeneren aan de hand van concepten*; materiële*, die betrekking heeft op het opstellen van en tussenkomen in proefnemingen omtrent het fenomeen in kwestie; en *sociale*, die volgt uit het feit dat de individuele onderzoeker zowel afhangt van als bijdraagt aan een of meerdere onderzoeksgemeenschappen die een deel van diens onderzoeksinteresses delen. De sociale context waarbinnen inzicht tot stand komt is essentieel om het als ‘wetenschappelijk’ te kunnen bestempelen, aangezien de persoonlijke ervaring van het begrijpen wordt omgezet in een overdraagbare, gedeelde ervaring doordat het individu deel uitmaakt van een of meerdere wetenschappelijke gemeenschappen. Mijn studie van de biologische praktijk laat mij toe om een filosofische beschouwing uit te werken over wetenschappelijk inzicht als een onafwendbaar pluralistische, persoonlijke ervaring die steunt op specifieke sociale, intellectuele en materiële omstandigheden. De focus op hedendaags biologisch onderzoek illustreert hoe de kijk van een bioloog op de wereld gevormd wordt door de vaardigheden en ‘commitments’ die hij verworven heeft door specifieke praktijkervaring.

Daarom stel ik een beeld voor van wetenschappelijk begrijpen als een cognitieve verworvenheid die volgt uit een doelmatige volbrenging van specifieke taken, zodat er tot verschillende interpretaties van eenzelfde fenomeen gekomen wordt door onderzoekers met verschillende epistemische vaardigheden en ‘research commitments’. Er is een verscheidenheid aan manieren waarop individuen hun capaciteit voor wetenschappelijk begrijpen verwerven, die daardoor verscheidene vormen kan aannemen als gevolg van de werktuigen aangewend tijdens het proces. In het geval van biologisch inzicht zal deze capaciteit afhangen van de benodigde achtergrondkennis die de individuele onderzoeker heeft opgedaan, tezamen met zijn of haar expertise in het hanteren van instrumenten, modellen en theorieën die het opstellen en toepassen van wetenschappelijke verklaringen toelaten. Het vereist ook de benodigde sociale vaardigheden om effectief te interageren met andere wetenschappers, om bijvoorbeeld de eigen onderzoeksresultaten te verspreiden en hun belang te beoordelen zowel in het licht van persoonlijke ervaring als langs de meetstaf van de relevante onderzoeksgemeenschap. Uit de toegankelijkheid van kennis volgt niet automatisch dat de gebruiker bewust is hoe deze kennis moet geinterpreteerd worden en hoe ze kan toegepast worden, dit bewustzijn moet daarentegen opgebouwd worden door jaren van training in specifieke wijzen van denken en handelen. Dat inzicht leidt mij tot de conclusie dat de capaciteit voor wetenschappelijk begrijpen een noodzakelijke voorwaarde is voor wetenschappelijke kennis; met andere woorden, in een wetenschappelijke context is het niet toelaatbaar dat een persoon zegt dat hij of zij een specifiek fenomeen ‘kent’, wanneer die persoon geen wetenschappelijke verklaring kan gebruiken om het fenomeen in kwestie te begrijpen.

Teneinde de wijdlopende draden van filosofisch, historisch en sociologisch materiaal tot een samenhangende analyse te weven, is dit proefschrift georganiseerd langs een lijn van toenemende complexiteit. Elk hoofdstuk bouwt verder op de voorgaande zodat de verschillende aspecten van mijn argumentatie een voor een belicht worden, waardoor ik hun filosofische betekenis en implicaties kan bespreken voor de eigenschappen en processen die het onderzoek aan Arabidopsis kenmerken.

Hoofdstuk 1 introduceert mijn onderwerp en de vragen die ik behandel in deze thesis, en wordt gevolgd door drie hoofdstukken die de filosofische, historische en methodologische basis van mijn argumentatie introduceren en rechtvaardigen. *Hoofdstuk 2* presenteert het filosofische kader voor mijn analyse, te beginnen met een overzicht van het hedendaagse filosofische debat over het onderscheid tussen verklaren en begrijpen. Mijn startpunt hier is Henk de Regts visie hoe de relatie tussen deze twee begrippen moet omschreven worden. Vervolgens bekijk ik de biologie van dichterbij en bespreek ik twee van haar kenmerken die ik van belang acht in een studie van het begrijpen: het pluralisme dat zowel theorieën als modellen kenmerkt, en wat ik de ‘tweeledige natuur’ van biologische kennis noem. Dit laatste omvat *theoretische kennis* (ondermeer datgene wat meestal gezien wordt als de inhoud van kennis, dat wat we beschouwen als feiten, theorieën, verklaringen en concepten over fenomenen die niet afhangen van specifieke procedures of interacties) en *belichaamde kennis* (die bestaat uit een bewustzijn van de benodigde handelswijzen en redeneervormen voor het uitvoeren van wetenschappelijk onderzoek). Binnen deze context geef ik een eerste definitie van het wetenschappelijke begrijpen als *de cognitieve verworvenheid die bereikbaar is voor individuele wetenschappers, afhankelijk van hun bekwaamheid in het coördineren van theoretische en belichaamde kennis over een specifiek fenomeen.* Doorheen *Hoofdstukken 5, 6* en *7* geef ik illustraties van deze definitie en diep ik haar verder uit, om haar tenslotte in *Hoofdstuk 8* een systematische behandeling te geven. Ik sluit het tweede hoofdstuk af met het voorstellen en bespreken van die elementen uit de sociale epistemologie die relevant zijn voor mijn onderzoek naar het vormen van individueel begrip door deelname aan een of meerdere onderzoeksgemeenschappen.

In *Hoofdstuk 3* schets ik de historische en sociale context van mijn onderzoek. Na een korte bespreking van ‘onderzoek aan modelorganismen’ in de biologie, belicht ik de oorsprong en ontwikkeling van de gemeenschap van onderzoekers die met Arabidopsis werken. Vervolgens bespreek ik de structuur en de doelstellingen die de gemeenschap heden ten dage heeft, en bestudeer ik nauwgezet welke sociale en wetenschappelijke rollen vervuld worden door twee onderzoeksgroepen waar het merendeel van mijn analyse zich op toespitst. De eerste is The Arabidopsis Information Resource [TAIR], ondergebracht bij het Carnegie Institute for Plant Biology in Stanford (VS).
Dit onderzoeksteam is belast met het opzetten van digitale databanken die (1) alle beschikbare gegevens over Arabidopsis toegankelijk zouden moeten maken voor geïnteresseerde onderzoekers, en (2) werktuigen zouden moeten aanreiken waarmee deze uiteenlopende gegevens kunnen samengesmeed worden, teneinde tot een geïntegreerd begrip van de biologie van Arabidopsis te komen. De tweede groep beheert het Nottingham Arabidopsis Stock Centre [NASC] vanuit de Universiteit van Nottingham (GB), waar verschillende types van Arabidopsis worden gecatalogeerd, gecultiveerd, bewaard en verspreid onder wetenschappers die ze gebruiken als modelorganisme.
Hoofdstuk 4 dient ter introductie en verdediging van mijn onderzoekswijze, die ik onder de noemer ‘geschiedenis, filosofie en sociologie van wetenschap en technologie’ plaats, in het licht van de filosofische en historische elementen die in de voorgaande hoofdstukken voorbij kwamen. Vertrekkende vanuit een discussie over de betekenis van het gebruik van case studies in het onderzoeken van algemene filosofische kwesties, bespreek ik de beweegredenen en bezorgdheden waar mijn onderzoek uit volgde, en breng ze in verband met de onderzoeksmethode die ik koos. Zoals ik in dit hoofdstuk benadruk, hoop ik dat dit onderzoek interessant is zowel voor de wetenschapsfilosofen en –historici als voor biologen met een filosofische inslag (zoals de vele wetenschappers die mij geholpen hebben door mij toegang te verlenen tot hun proefruimtes, personeel en gedachten over onderzoek aan Arabidopsis). In die geest neem ik de kijk op ‘complementaire wetenschap’ over van Hasok Chang zoals hij die recent introduceerde, en breid ik deze uit door te stellen dat wetenschapsgeschiedenis en –filosofie kunnen en moeten trachten ons begrip van de wereld te verbeteren, zodat ze een goed deel van de doelstellingen en waarden van de natuur- en sociale wetenschappen delen.

Hoofdstukken 5, 6 en *7* vormen de centrale argumentatie van dit proefschrift. Elk bekijkt een ander aspect van biologisch inzicht van naderbij, geïllustreerd aan de hand van onderzoek aan Arabidopsis. *Hoofdstukken 5* en *6* staan uitgebreid stil bij de rol in het begrijpen van biologische fenomenen dewelke gespeeld wordt door twee types van epistemische vaardigheden, die ik *theoretisch* en *uitvoerend* noem. *Hoofdstuk 5* richt zich op de onderzoekspraktijk binnen TAIR, die bestaat uit het vervaardigen van databanken die zoveel mogelijk beschikbare gegevens over Arabidopsis kunnen visualiseren. Ik toon aan dat zowel het ontwerp als het gebruik van dergelijke gegevensbanken noodzakelijkerwijze refereert aan een theoretisch kader, in casu het netwerk van concepten dat de TAIR onderzoekers gebruiken om de gegevens te structureren. Ik betoog dat dit netwerk van concepten dat biologen duiden met het woord 'bio-ontologie' geen neutraal hulpmiddel is voor het verpreiden van gegevens uit experimenten. Het representeert veeleer een nieuw type van theorie in de biologie, die biologische betekenis kleeft aan gegevens tegelijk met het beschikbaar stellen ervan. De bespreking van bio-ontologieën leidt mij tot het onderscheiden van de twee genoemde types van vaardigheden bij de onderzoekers, waardoor ik de schijnwerper kan richten op het feit dat Arabidopsisonderzoekers die verschillende vaardigheden gebruiken tot een verschillend begrip kunnen komen van eenzelfde fenomeen. Deze bewering wordt uitgewerkt in *Hoofdstuk 6,* waarin ik de werkzaamheden bespreek van NASC onderzoekers bij het produceren van Arabidopsis specimens verspreiding onder laboratoria over de gehele wereld. Ik stel dat zowel TAIR als NASC met succes modellen maken van veldstammen van Arabidopsis: in het eerste geval bestaan deze modellen uit beelden van de biologie van de plant, in het tweede geval uit fysieke organismen met eigenschappen die gekozen en aangepast zijn om aan de eisen van de proefnemers te voldoen. De belangrijke verschillen in de benodigde vaardigheden om te werken met deze twee soorten van modellen komen naar voren bij een vergelijking van de vervaardiging en het gebruik van beide. Door een analyse van de wijze waarop deze modellen geabstraheerd worden uit respectievelijk gegevensverzamelingen over en daadwerkelijke exemplaren van Arabidopsis, kan ik uitweiden over de manier waarop de onderzoeker een samenspel van theoretische en uitvoerende vaardigheden kan gebruiken om met deze modellen zijn of haar begrip van de biologie te vergroten.

Hoofdstuk 7 voegt een essentieel niveau van complexiteit toe aan de analyse tot dusver, door de sociale dynamiek rondom het Arabidopsisonderzoek te bespreken. Ik typeer de Arabidopsisgemeenschap als een geval van gecentraliseerde Big Science, en analyseer het effect dat deze institutionele structuur heeft op het verwerven en verspreiden van wetenschappelijk begrip in deze gemeenschap. In het bijzonder maak ik een lijst van sociale vaardigheden die ik noodzakelijk acht om de biologie van Arabidopsis te begrijpen, in zoverre dat ze toegang verlenen tot de theoretische en uitvoerende vaardigheden nodig voor het uitvoeren en delen van onderzoek met collega's.
Ik toon ook aan dat aandacht voor deze vaardigheden complementair kan zijn voor bestaande benaderingen van sociale epistemologie in de wetenschap, waarbij het 'kritisch-contructieve empirisme' van Helen Longino het meest voor de hand ligt.

Hoofdstuk 8 tenslotte presenteert de epistemologische conclusies die ik trek uit mijn analyse van hoe het handig gebruikmaken van theorieën, modellen en door de gemeenschap gedeelde bronnen de Arabidopsiswetenschappers toelaat om hun begrip van de plantkunde te vergroten. Ik duid op de sterke band tussen de vaardigheden van onderzoekers en de verbintenissen die ze aangaan, voor zover het handelingen en overtuigingen betreft waarbij hun kennis en vaardigheden een rol spelen. Daarna beschouw ik drie grote types van biologisch inzicht in een fenomeen die alle drie volgen uit specifieke combinaties van vaardigheden en ‘commitments’ die onderzoekers gebruiken om hun theoretische en belichaamde kennis van dat fenomeen met elkaar in overeenstemming te brengen: *theoretisch, belichaamd* en *geïntegreerd* begrijpen. Het bestuderen van de epistemische rol van geïntegreerd begrip blijkt vooral interessant in verband met de pogingen die momenteel ondernomen worden meerdere biologische onderzoeksgebieden te fuseren of integreren, om tot een ‘volledig’ begrijpen van een organisme te komen. Tot besluit haal ik mijn methodologische insteek weer naar voren, in verband met complementaire wetenschap zoals voorgesteld in *Hoofdstuk 4*, en ik bespreek een aantal manieren waarop mijn onderzoek over begrip een impact kan hebben op de dagelijkse biologische praktijk.